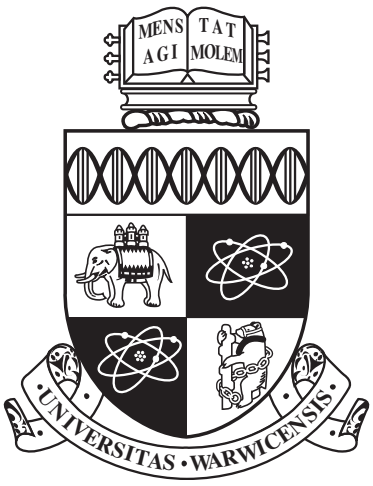


The Role of Histone Acetyltransferases in Plant Immunity

by

Stephanie Jayne Thorley Kancy

Thesis

Submitted to The University of Warwick

for the degree of

Doctor of Philosophy

School of Life Sciences

December 2017

THE UNIVERSITY OF
WARWICK

Contents

List of Tables **vi**

List of Figures **vii**

Declarations **x**

Acknowledgments **xi**

Abstract **xiii**

Abbreviations **xiv**

Chapter 1 Introduction **1**

1.1 Motivation: achieving and maintaining food security 1
1.2 *A. thaliana* and *P. syringae* as a model pathosystem 3
1.3 Immune responses in *A. thaliana* 4
1.3.1 PAMP-triggered immunity (PTI) 4
1.3.2 Effector-triggered susceptibility (ETS) 6
1.3.3 Effector-triggered immunity (ETI) 6
1.3.4 Hormone signalling in defence 9
1.3.5 Defence and growth trade-off 9
1.3.6 Transcriptional reprogramming during infection 10
1.4 Regulation of gene transcription in plants 11
1.4.1 Chromatin remodelling 11
1.4.2 Histone modifications 12
1.5 Histone acetylation 14
1.5.1 *Arabidopsis* histone acetyltransferases 14
1.6 Aims of this study 19

Chapter 2 Material and Methods **21**
2.1 Plants used in this study . . . 21
2.1.1 Plant growth conditions . . . 21
2.1.2 *Arabidopsis* mutant lines . . . 22
2.1.3 Generation of *Arabidopsis* crosses . . . 22
2.1.4 Stable transformation of *Arabidopsis* . . . 23
2.2 Plant pathogen material . . . 23
2.2.1 *P. syringae* material and growth . . . 23
2.2.2 *B. cinerea* material and growth . . . 23
2.3 Plant pathology assays . . . 24
2.3.1 Flg22 peptide elicitation . . . 24
2.3.2 *P. syringae* infection . . . 24
2.3.3 Quantification of bacterial growth in infected leaves . . . 25
2.3.4 Ion leakage measurement . . . 25
2.3.5 *B. cinerea* infection . . . 26
2.4 Plant development assays . . . 26
2.4.1 Leaf surface area and fresh weight measurement . . . 26
2.4.2 Primary root measurement . . . 26
2.4.3 Root apical meristem measurement . . . 27
2.5 Molecular biology methods . . . 27
2.5.1 DNA methods . . . 27
2.5.2 Polymerase chain reaction (PCR) . . . 27
2.5.3 Gel electrophoresis . . . 28
2.5.4 DNA extraction from agarose gel . . . 28
2.5.5 Gateway cloning . . . 29
2.5.6 DNA sequencing . . . 31
2.5.7 RNA methods . . . 31
2.6 Agilent microarrays . . . 32
2.6.1 RNA extraction . . . 32
2.6.2 Amplification and labelling . . . 32
2.6.3 Hybridisation, washing and scanning . . . 33
2.6.4 Data normalisation and differential expression . . . 33
2.6.5 GO term and motif enrichment . . . 33
2.7 Homology modelling . . . 34
2.7.1 Template selection and sequence alignment . . . 34
2.7.2 Model construction and evaluation . . . 34

2.8 Compound library preparation 35
2.9 *In silico* docking 36

Chapter 3 HAM2 is a Moderator of Plant Immunity and Growth 37
3.1 Introduction 37
3.1.1 Aims of this chapter 42
3.2 HAM2 is a negative regulator of immunity 43
3.2.1 *ham2* mutants show enhanced resistance to *P. syringae* infection 43
3.2.2 Microarrays following flg22 treatment 44
3.2.3 Effector-triggered immunity is not compromised in *ham2* mutants 59
3.2.4 Defence against *B. cinerea* is not compromised in *ham2* mutants 60
3.2.5 Flg22 growth inhibition assay 61
3.3 HAM2 is a negative regulator of plant growth 63
3.3.1 Adult *ham2* mutants have enhanced leaf surface area and fresh weight 63
3.3.2 *ham2* seedlings have enhanced root length 64
3.3.3 *ham2* seedlings have enhanced root apical meristem size . . . 64
3.4 Discussion 66
3.4.1 The function of HAM2 in defence responses 66
3.4.2 HAM2 is a negative regulator of vegetative growth 70
3.4.3 Further work 71
3.4.4 HAM2 as a novel target in an agricultural context 71

Chapter 4 Identification of Inhibitors Targeting HAM2 73
4.1 Introduction 73
4.1.1 Development of inhibitors targeting MYST HATs 74
4.1.2 *In silico* docking for inhibitor identification 74
4.1.3 *In silico* docking programs 76
4.1.4 Homology modelling 76
4.1.5 Process of work in this chapter 80
4.2 Homology modelling of AtHAM2, BnHAM2, SlHAM2 and AtHAG1 . 82
4.2.1 Identification of AtHAM2 homologues 82
4.2.2 Template identification 84
4.2.3 Sequence alignment 87
4.2.4 Model construction and quality assessment 88
4.2.5 Structural comparison of AtHAM2 and HsKAT8 91

4.3 Compound database preparation and filtering . . . 94
4.4 *In silico* docking to identify novel inhibitors . . . 95
4.4.1 Validation of AutoDock Vina docking algorithm . . . 95
4.4.2 Screen set-up . . . 96
4.4.3 Identifying AtHAM2-selective compounds from screen results . 99
4.4.4 Docking AtHAM2-selective compounds against BnHAM2, SlHAM2 and AtHAG1 . . . 100
4.4.5 Comparing Ac-CoA interactions with test compound interactions using a novel program- Sifter . . . 102
4.4.6 Visual inspection of top-ranking candidate inhibitors . . . 103
4.4.7 Interactions of candidate inhibitors with AtHAM2 binding pocket . . . 106
4.5 Initial *in planta* testing . . . 109
4.6 Discussion . . . 111
4.6.1 MYST-family HATs are highly conserved . . . 111
4.6.2 Molecular modelling of AtHAM2, BnHAM2, SlHAM2 and AtHAG1 . . . 111
4.6.3 Identification of inhibitors . . . 112
4.6.4 Limitations and further work . . . 113

Chapter 5 Investigating the Role of HATs in Effector-Triggered Immunity 115
5.1 Introduction . . . 115
5.1.1 HAG1 is a positive regulator of basal defence . . . 116
5.1.2 Effector-triggered immunity . . . 118
5.1.3 Aims of this chapter . . . 120
5.2 Assessing AvrRpt2-induced responses in *hat* mutants . . . 120
5.2.1 AvrRpt2-induced ion leakage is attenuated in *hac4*, *hac5*, *ham1* and *hag1* mutants . . . 120
5.2.2 AvrRpt2-immunity is not affected in *hac4*, *hac5* and *ham1* mutants, but is in *hag1-6* . . . 124
5.3 The role of HAG1 in ETI . . . 126
5.3.1 AvrRpt2, AvrRpm1 and AvrPphB-induced ion leakage is attenuated in *hag1-6* . . . 126
5.3.2 AvrRpt2, AvrPphB and AvrRps4-immunity is compromised in *hag1-6* . . . 127

5.4 Transcriptomic analysis of *hag1-6* ETI responses 129
5.4.1 ETI marker gene expression in *hag1-6* 129
5.4.2 Microarray analysis of Col-0 and *hag1-6* in response to *Pto* (EV) and *Pto* (*avrRpt2*) 131
5.5 Discussion 145
5.5.1 AvrRpt2-induced responses of *hac4*, *hac5* and *ham1* 145
5.5.2 ETI responses in *hag1-6* 146
5.5.3 Limitations and further work 147

Chapter 6 General Discussion **148**
6.1 HAM2 moderates plant immunity and growth 149
6.2 Targeting chromatin modulators for crop improvement 152
6.3 HAG1 plays a central role in ETI 153
6.4 General outlook 155

List of Tables

2.1 *Arabidopsis thaliana* mutant lines used in this study. 22
2.2 *Pseudomonas syringae* pv. *tomato* strains used in this study. 24
2.3 Thermal cycling conditions used for genotyping PCR reactions. . . . 28
2.4 Thermal cycling conditions used for attB PCR reactions. 30
2.5 Thermal cycling conditions used for qPCR amplification reactions. . . 32
2.6 Protein Data Bank (PDB) structures used in this study. 34

3.1 GO terms significantly enriched in the 114 flg22-responsive genes with enhanced expression in *ham2-75*. 53
3.2 Over-represented known transcription factor binding motifs in the 114 flg22-responsive genes with enhanced expression in *ham2-75* than Col-0 after flg22 treatment. 54
3.3 Sixteen of the 114 flg22-responsive genes with enhanced expression in *ham2-75* that contain the WRKY18 motif and are targeted by WRKY18 after treatment with 1 M flg22. 56

4.1 Percent identity matrix of HAM2 homologues. 84
4.2 Identification of structural templates for homology modelling. 86
4.3 Top 10 Sifter overlap scorers 103
4.4 Structure of the candidate lead compounds. 105

5.1 The number of differentially expressed genes in microarray studies utilising AvrRpt2. 134

List of Figures

1.1 The "zig-zag" model describes quantitative outputs of plant-pathogen interactions. 5
1.2 Signalling events and consequences following pathogen perception. . . 8
1.3 Histone modifying enzymes and chromatin remodelling complexes determine chromatin structure. 13
1.4 Acetylation and de-acetylation of lysine residues. 15
1.5 Phylogenetic tree and domain architecture of the *Arabidopsis* HATs. . 16

3.1 HAT mutant lines *ham2-75* and *hag1-6* are more resistant and more susceptible to *Pto* DC3000 respectively. 39
3.2 Expression of *HAM2* (AT5G09740) in response to infection with *B. cinerea*, *Pto* DC3000 and *Pto* DC3000*hrpA*-. 41
3.3 Two *ham2* mutant lines are resistant to *Pto* DC3000. 43
3.4 Overall microarray intensities before and after quantile normalisation. 45
3.5 Heat map representation of differential gene expression in Col-0 and *ham2-75* prior to and following flg22 elicitation. 46
3.6 Identified flg22 up-regulated genes in Col-0 and *ham2-75*. 50
3.7 The difference in expression values of genes up-regulated after exposure to flg22 in Col-0 and *ham2-75*. 52
3.8 Expression of *HAM2* (AT5G09740) and WRKY18 binding motif containing genes in response to infection with *Pto* DC3000 and *Pto* DC3000*hrpA*-. 58
3.9 AvrRpt2-induced HR is not affected in *ham2* mutant lines. 60
3.10 Defences against the necrotrophic fungus *B. cinerea* are not compromised in *ham2* mutant lines. 61
3.11 *ham2* mutants retain an enhanced fresh weight phenotype under flg22-induced stress. 62

3.12 *ham2-71* and *ham2-75* mutant lines display an enhanced leaf surface area and fresh weight phenotype. 63
3.13 *ham2* mutant lines display an enhanced root length phenotype. . . . 64
3.14 Cell number in the root apical meristem is enhanced in *ham2* mutants. 65
3.15 Sequence alignment of *Arabidopsis* HAM1 and HAM2. 67

4.1 The "safe" and "twilight" zones of homology modelling. 78
4.2 Pictorial depiction of the work flow in this chapter 81
4.3 The predicted domain structures of HAM2 homologues depicted in cartoon representation. 83
4.4 Sequence alignment of the HAT domains of HsKAT8 and AtHAM2. . 87
4.5 Evaluation of AtHAM2 homology model quality. 90
4.6 Structural comparison of the AtHAM2 model and HsKAT8. 93
4.7 Residue and electrostatic surface potential conservation between AtHAM2 and HsKAT8. 94
4.8 Autodock Vina validation: docking of Ac-CoA into HsKAT8. 96
4.9 Grid-box position over the Ac-CoA binding site of AtHAM2 and 4DNC (HsKAT8). 98
4.10 A compound that docks with a similarly high docking score into AtHAM2 and HsKAT8. 100
4.11 Docking scores difference between HsKAT8 and AtHAM2 of all compounds screened. 101
4.12 2D interaction plots between AtHAM2 and candidate lead compounds. 108
4.13 Cell number in the root apical meristem is enhanced in Col-0 when treated with NU9056. 110

5.1 T-DNA insertion position and morphological phenotype of the *hag1-6* line. 117
5.2 *hat* mutants *hac4*, *hac5*, *ham1* and *hag1* have compromised AvrRpt2-induced HR. 122
5.3 T-DNA insertion positions and morphological phenotypes of *hac4-21*, *hac5-78* and *ham1-96* lines. 123
5.4 *Pto* (EV) and *Pto* (*avrRpt2)* growth is affected in *hag1*, but is not affected in *hac4*, *hac5* and *ham1* mutants. 125
5.5 AvrRpt2-, AvrRpm1- and AvrPphB-induced HR is compromised in *hag1-6*. 127

5.6 *hag1-6* is more susceptible to *Pto* (EV), *Pto* (*avrRpt2*), *Pto* (*avrPphB*) and *Pto* (*avrRps4*). 128
5.7 ETI marker gene expression is attenuated in *hag1-6* compared to Col-0. 130
5.8 Overall microarray intensities before and after quantile normalisation. 133
5.9 Heat map representation of differential gene expression between Col-0 and *hag1-6* after inoculation with mock (10 mM $MgCl_2$), *Pto* (EV) and *Pto* (*avrRpt2*). 136
5.10 Differentially expressed genes between Col-0 and *hag1-6* under mock conditions. 138
5.11 Differential expression between the Col-0 and *hag1-6* transcriptome when challenged with *Pto* (EV) and *Pto* (*avrRpt2*). 141
5.12 Venn diagrams of differentially expressed genes between Col-0 and *hag1-6* after infection with *Pto* (EV) or *Pto* (*avrRpt2*) compared to mock. 142
5.13 BINGO visualisation of the genes differentially expressed in Col-0 but not *hag1-6* upon infection with *Pto* (*avrRpt2*). 144

6.1 A summary of the *ham2* and *hag1* phenotypes presented in this thesis. 149
6.2 The role of HAM2 in basal defence responses. 151
6.3 HAG1 plays a central role in ETI following Avr protein perception. . 154

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. No part of the work submitted here has also been submitted in any previous application for any degree. The work presented in this thesis was undertaken by myself except where otherwise stated.

Acknowledgments

Firstly, I would like to sincerely thank Dr Vardis Ntoukakis for welcoming me into his group and for his supervision during the past 3 years. Thank you for the opportunity to work on such an interesting and varied project and for the guidance throughout. I'd also like to thank my secondary supervisor, Professor Katherine Denby, for her contributions to this work and also my advisory panel, Dr Miriam Gifford and Dr Patrick Shafer, for their suggestions during our panel meetings.

Thank you to Professor Richard Napier and his group members for much-appreciated encouragement and guidance during my adventures into *in silico* modelling. My special thanks go to Dr Charo I. del Genio and Dr Vesi V. Uzunova for their abounding knowledge, patience and kindness. Thank you also to Dr Mussa Quareshy for his enthusiastic sharing of very useful computational knowledge. Thank you also to Dr Sascha Ott for his transcriptional analysis advise and for his excellent pad-holding abilities during boxing sessions. I also very much appreciate the help and kindness of Dr Paul Brown for his assistance with the HPC cluster.

My most special and deepest thanks go to Dr Sophie Piquerez, who made the majority of the work in this thesis possible. I can't thank her enough for all her time, patience, wisdom, humour, proof-reading abilities and encouragement throughout.

To all of the C030 "golden oldies", thank you for the memories: Claire, Christine, Sarah, Manos, Jack, Chrysa (I owe you lots of delicious food when it's your turn!). Special thank you to Alonso (Mr Tumnus) for being so supportive and for

being Alonso. To everyone from my new office (Elena (gracias por ensearme LaTeX!), Kate, Hannah, Olga, Denise and Matt): it was a joy sharing an office with such lovely people!

A huge thank you to my wonderful friends and family who have been a constant source of support and inspiration: my brother Scott, Kerry, Izzy (kaka!), Grem Master and Alex, thank you all for being so great. Thank you to my housemates old and new: Jim, Illiana, Ellie and Te Anne - it was short, but ever so sweet!

Finally, I dedicate this thesis to my mum, Debbie. I couldn't have got to this point without her endless encouragement, support and love. Thank you so much for everything, mum.

Abstract

Activation of plant defence responses requires significant transcriptional reprogramming to mount an effective response to pathogens. This response must be finely balanced with growth and development processes to ensure optimal allocation of cellular resources. A fundamental mechanism of gene expression regulation is covalent modification of histones. Histone acetyltransferases (HATs) and histone deacetylases (HDACs) antagonistically control the acetylation levels of histones at specific genomic loci to ultimately affect gene expression. This thesis focuses on histone acetylation as a mechanism by which plants mount an effective immune response.

In Chapter 3, a reverse genetic screen of *Arabidopsis* HAT mutants is presented where a negative regulator (HAM2) of defence against the plant pathogen *Pto* DC3000 was identified. Whilst mutants of the negative regulator (*ham2*) demonstrate enhanced resistance to *P. syringae*, their susceptibility to the necrotrophic pathogen *B. cinerea* is unchanged. Alongside the immunity phenotype, *ham2* plants exhibit increased adult leaf surface area, fresh weight and root length.

Since *ham2* is the only known *Arabidopsis* mutant with increased immunity and growth, it represents a promising target in an agricultural context. In Chapter 4, homology models of *A. thaliana*, *B. napus* and *S. lycopersicum* HAM2 proteins were created, supported by a series of cheminformatics and *in silico* docking methods, to identify chemical inhibitors for future agricultural applications.

Finally, the role of *Arabidopsis* HATs in effector-triggered immunity was investigated in Chapter 5. Here, HAG1 was identified as a key positive regulator of effector-triggered responses. Overall, this thesis contributes to our understanding of the role of HAM2 and HAG1 histone acetyltransferases in plant immunity, and presents HAM2 as a novel target in an agricultural context.

Abbreviations

Ac-CoA	Acetyl Coenzyme A
Avr	Avirulence
BLAST	Basic local alignment search tool
bp	Base pairs
cDNA	Complementary DNA
CFU	Colony forming units
ChIP	Chromatin immunoprecipitation
DEG	Differentially expressed gene
DNA	Deoxyribonucleic acid
DOPE	Discrete optimised protein energy
dpi	Days post infection
ETI	Effector triggered immunity
EV	Empty vector
FDR	False discovery rate
GEO	Gene expression omnibus
GO	Gene ontology
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
hpi	Hours post infection
HR	Hypersensitive response
JA	Jasmonic acid
LIMMA	Linear models for microarray data
molpdf	Molecular potential density function

MS	Murashige & Skoog medium
OD	Optical density
PAMP	Pathogen-associated molecular pattern
PCR	Polymerase chain reaction
PDB	Protein data bank
PTI	PAMP-triggered immunity
pv.	Pathovar
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SA	Salicylic acid
SAR	Systemic acquired resistance
T-DNA	Transfer DNA
TAIR	The Arabidopsis information resource
TF	Transcription factor

Chapter 1

Introduction

1.1 Motivation: achieving and maintaining food security

One of the biggest challenges facing the global community is securing sustainable food sources for the exponentially growing human population. In March 2012, the world population reached 7 billion people. It is estimated that by 2050, this figure will have reached 9.5 billion (UN, 2013); this is a rise of 2.5 billion people in just 38 years. In order to maintain sufficient calories per capita, concerted efforts must be made to sustainably increase crop production yields. Increasing the amount of land used for agriculture is not a viable solution, since this would lead to adverse impacts on areas with ecological importance, such as tropical rain forests and habitats where predators of crop pests reside (Gibbs et al., 2010). Not only is the population rising, but dietary habits are changing such that consumption of meat and milk products is rising, increasing the demand of grains for livestock feed (Popp et al., 2013). Climate change is predicted to exacerbate agricultural challenges by enhancing abiotic stress and by creating conditions in which novel plant diseases may emerge and become epidemic (Boyd et al., 2013).

Economically important crops are exposed to a variety of pathogens such as viruses, bacteria, oomycetes and fungi that have different lifestyle modes and infection strategies. It is estimated that one quarter of crops are lost pre-harvest worldwide due to pests and diseases. The colonisation and deterioration of crops by pathogenic organisms therefore presents a major constraint to food security (Boyd et al., 2013). As a prominent example, the oomycete *Phytophthora infestans* was the causative

agent of late blight in tomato and potato, which caused the deaths of 1.25 million people during the Irish potato famine in 1845 (Agrios, 2005). The same pathogen poses contemporary issues also: yearly global losses due to potato late blight are estimated to be €4.8 billion (Haverkort et al., 2008). Another example includes *Ralstonia solanacearum* which is considered to be the most destructive bacterial phytopathogen worldwide due to its geographical distribution and wide host range. For potatoes alone, *Ralstonia solanacearum* is estimated to cause an annual loss of US $1 billion worldwide. The bacteria first colonises the root cortex, followed by invasion of the xylem where it multiplies to very high population levels causing wilting and plant death (Mansfield et al., 2012).

Despite these agricultural challenges, vast improvements in productivity have been made over the past 50 years. This has been achieved through technological advances such as the development of chemicals like pesticides (Aktar et al., 2009). In a similar scenario to the emergence of antibiotic-resistant microbes, plant pests and pathogens can also evolve mechanisms of resistance to agrichemicals. In fact, a study by Roush and Tabashnik (2012) discovered that over 400 insects or mites had developed resistance to one or more pesticides. Taking into account the other disadvantages of pesticides, such as effects on non-target organisms and contamination of ground and surface water (Aktar et al., 2009), novel strategies for control are required to improve yield and minimise cases of resistance.

One strategy, which has been pursued through many decades of research, is to study plant abiotic and biotic stress responses. The fundamental aim of this research is to enhance plant resistance through modification of responses to stimuli. In the field of biotic stress, there have already been successful cases of conferred resistance through the interfamily transfer of defence receptor genes (Lu et al., 2015; Schwessinger et al., 2015; Holton et al., 2015). For example, the EF-Tu receptor (EFR), which was previously restricted to the dicot *Brassicaceae* family, has been transferred to *Solanaceae* (Lacombe et al., 2010) and the monocot *Triticum* family (Schoonbeek et al., 2015) to confer broad-spectrum bacterial resistance. Similarly the nucleotide binding-leucine rich repeat (NB-LRR) genes, *RPS4* and *RRS1*, were transferred to multiple *Brassicaceae* and *Solanaceae* species from *Arabidopsis thaliana* to confer resistance to several bacterial families (Narusaka et al., 2013; Maekawa et al., 2012). The success of these studies, and many others like them, comes from the extensive knowledge of plant immunity. This has been cultivated

through analysis of model plant-pathogen systems, such as *Arabidopsis thaliana* and the bacterium *Pseudomonas syringae*.

1.2 *A. thaliana* and *P. syringae* as a model pathosystem

In all aspects of plant research, *Arabidopsis thaliana* has been indispensable as a model organism. It possesses one of the smallest plant genomes (115.5 Mbp across 5 chromosomes) and was the fourth genome sequence to be published (Kaul et al., 2000). A vast library of mutants, created through EMS mutagenesis or T-DNA insertion, has enabled the in-depth analysis of specific genes in particular processes through exposing knock-out or over-expressor mutant lines to a condition of interest. In the context of immunity, *Arabidopsis thaliana* can be infected by oomycetes (e.g. *Hyaloperonospora arabidopsis*), bacteria (e.g. *Pseudomonas syringae*), fungi (e.g. *Botrytis cinerea*) and viruses (e.g. cauliflower mosaic virus) and responds to infections in a similar way to other higher plants.

Pseudomonas syringae is a Gram-negative rod-shaped bacterium which infects a wide range of plant species. It damages economically important crops, including tomatoes where it causes bacterial speck disease leading to significant yield and financial losses (Scofield et al., 1996). Several pathovars of *Pseudomonas syringae*, such as *tomato*, *pisi* and *maculicola*, were the first pathogens discovered to infect *Arabidopsis* in a laboratory setting (Katagiri et al., 2002). Best described as a hemibiotrophic pathogen, *P. syringae* displays both biotrophic and necrotrophic lifestyle modes; initially the pathogen enters the leaf tissue through stomata or wounds and lives as a pathogenic endophyte. Under high humidity conditions, the bacteria multiplies aggressively in the apoplastic space. At the later stages of colonisation, host cells die and leaf tissue becomes necrotic causing symptoms such as leaf spots and stem cankers (Jin et al., 2003; Xin and He, 2013). The importance of the *Arabidopsis-P. syringae* pathosystem became established when investigating gene-for-gene relationships describing the recognition of pathogen avirulence (*avr*) genes by host resistance (*R*) genes.

1.3 Immune responses in *A. thaliana*

Unlike animals, plants lack an adaptive immune system and therefore rely on innate immunity. However, the majority of plants are resistant to complete microbial species through the evolution of a sophisticated and multi-layered defence response (Heath, 2000). A simplified representation of the plant immune system was proposed in the zig-zag model by Jones and Dangl (2006), which describes how plants alter the amplitude of defence in response to pathogen tactics (Fig. 1.1). The first passive layer of defence consists of structural or chemical barriers, such as the epidermal waxy cuticle. If a pathogen breaches these preformed defences, a breadth of responses can occur. The first active layer of defence is through extracellular perception of conserved pathogen-associated molecular patterns (PAMPs) by membrane-associated pattern recognition receptors (PRRs) (Dodds and Rathjen, 2010). This triggers PAMP-triggered immunity (PTI), which includes reactive oxygen species (ROS) bursts, activation of kinase cascades, calcium ion influx, defence-related gene reprogramming and production of antimicrobial compounds. Virulent pathogens have evolved the ability to deploy proteins, known as effectors, into the host cell or apoplast to suppress PTI responses and aid disease progression. This manipulation of host processes is known as effector-triggered susceptibility (ETS). In turn, plants have evolved to directly or indirectly recognise pathogen effectors through NB-LRR proteins, initiating effector-triggered immunity (ETI). The pathogen may then evolve to discard recognised effectors, mutate in order to avoid recognition or gain more effectors to suppress ETI to continue the host-pathogen arms race.

1.3.1 PAMP-triggered immunity (PTI)

The first layer of inducible defence, known as PTI, is mediated by PRRs which perceive PAMPs at the cell surface. PAMPs are invariant molecules associated with particular taxonomic classes, and are very difficult for the pathogen to modify or discard. PTI therefore restricts the growth of the majority of potential pathogens encountered by plants (Boller and Felix, 2009). Many PRRs are transmembrane leucine-rich repeat receptor kinases (LRR-RKs) which recognise conserved epitopes of molecules that are necessary for microbial fitness. One exemplary elicitor of PTI in *Arabidopsis* is flg22, a 22-amino acid fragment of bacterial flagellin which is necessary for its locomotion (Zipfel and Felix, 2005). The *Arabidopsis* PRR that perceives the flg22 epitope is FLS2 (Gómez-Gómez and Boller, 2000), which binds flg22 with its LRR ectodomain. This binding event allows the co-receptor BAK1

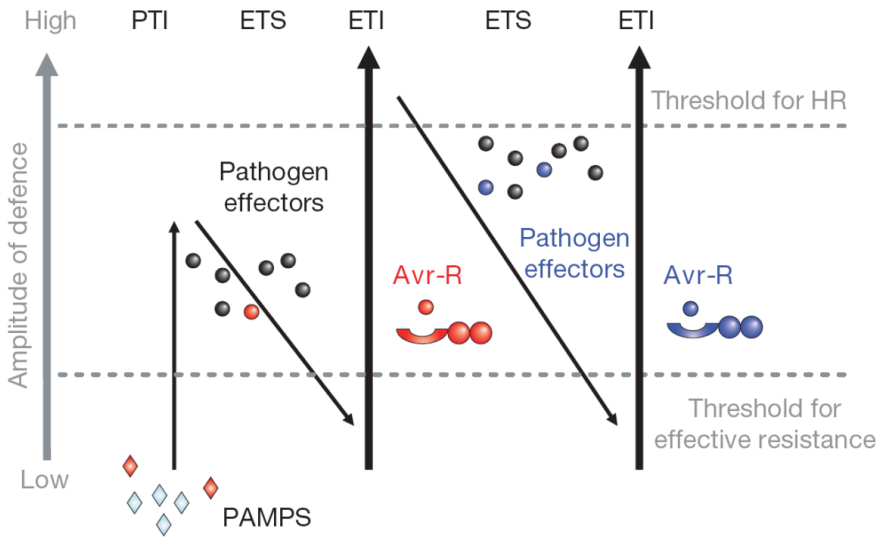


Figure 1.1: The "zig-zag" model describes quantitative outputs of plant-pathogen interactions. The first inducible phase of defence is activated when a PAMP (pathogen-associated molecule patterns, red diamonds) is detected, stimulating PTI (PAMP-triggered immunity). Successful pathogens deploy effector proteins (black circles) to dampen host PTI responses and aid pathogenesis which is known as ETS (effector-triggered susceptibility). A host plant may perceive a pathogen effector (Avirulence (Avr) protein, red circle) via R (Resistance) proteins, activating ETI (effector-triggered immunity). ETI is an amplified version of PTI which culminates in cell death by the HR (hypersensitive response). To overcome ETI, pathogens may shed recognised effectors (red circles) or gain new effector variants that attenuate ETI signalling (blue circles). In turn, new plant *R* gene alleles that detect the new effector variants will be selected for, again resulting in ETI and HR. The plant-pathogen arms race continues in this way. Figure taken from Jones and Dangl (2006).

(also a LRR-RK) to form a complex with FLS2 within 15 seconds of elicitation (Schulze et al., 2010). The formation of this heterodimer brings the cytoplasmic kinase domains into contact, initiating downstream defence signalling (Boller and Felix, 2009). These findings suggest that PRRs act as cell-surface "sentinels" which can perceive PAMPs and initiate a rapid response. The importance of FLS2 is underlined by the finding that mutation of *Arabidopsis FLS2* causes susceptibility to *Pseudomonas syringae* pv. *tomato* (*Pto*) DC3000 through the reduced activation of PTI (Zipfel et al., 2004). Other examples of well-characterised PRRs are EFR (EF-Tu receptor), which recognises the n-terminal 18 amino acids of bacterial elongation factor Tu (Kunze et al., 2004) and CERK1 (Chitin elicitor receptor kinase 1) which was identified in *Arabidopsis* for its role in perception of fungal chitin (Miya

et al., 2007).

The cellular events following PRR activation include Ca^2+ ion influx and the production of reactive oxygen species (Macho and Zipfel, 2014). Activation of calcium-dependent protein kinases (CDPK) and mitogen-activated protein kinases (MAPK) phosphorylation cascades (such as the MEKK1, MKK4/MKK5 and MPK3/MPK6 cascade (Asai et al., 2002)) transduce the PRR activation signal to downstream targets (as depicted in Fig. 1.2). The cellular outcomes resulting from these signalling cascades include, but are not limited to, callose deposition, stomatal closure, hormone biosynthesis and substantial transcriptional changes. These transcriptional changes lead to the expression of numerous pathogenesis-related (*PR*) genes via WRKY and other transcription factors families (TFs) (Asai et al., 2002; Zipfel et al., 2004; Kunze et al., 2004; Navarro et al., 2004).

1.3.2 Effector-triggered susceptibility (ETS)

The PTI responses described above are effective against the majority of potential pathogens. Virulent pathogens, such as *Pto* DC3000, have adapted to attenuate PTI responses by deploying an estimated 28 effector proteins directly into the plant cell via the needle-like type 3 secretion system (T3SS) (Cunnac et al., 2009). Effectors manipulate host signalling and metabolic processes in a variety of ways to facilitate the infection process (Jones and Dangl, 2006). For example, the *Pto* DC3000 effectors AvrPto and AvrPtoB can form complexes with the cytoplasmic kinase domain of BAK1, preventing the formation of the FLS2-BAK1 heterodimer which abrogates FLS2-dependent PTI signalling (Shan et al., 2008). Also, AvrPtoB has also been shown to ubiquitinate FLS2, BAK1 and CERK1, targeting them for proteasomal degradation (Göhre et al., 2008; Shan et al., 2008; Gimenez-Ibanez et al., 2009). Another *P. syringae* effector, HopAI1, was found to dephosphorylate MPK3 and MPK6, causing the suppression of both cell wall reinforcement and transcriptional activation of specific PAMP responsive genes (Zhang et al., 2007). Effectors are deployed by a wide range of phytopathogens to overcome host PTI defences and promote pathogenesis.

1.3.3 Effector-triggered immunity (ETI)

Plants have evolved the ability to respond to the activity of effectors through resistance (R) proteins. R proteins interact directly with effectors, or indirectly by

guarding the targets of effectors, activating the second wave of defence responses known as effector-triggered immunity (ETI) (Jones and Dangl, 2006). If the matching R protein is absent, the pathogen effector is able to exert its virulence function and the interaction is compatible. An incompatible interaction occurs when an effector is detected by its cognate R protein, in which case the effector is known as an avirulence (Avr) protein (Flor, 1971). The majority of R proteins are polymorphic NB-LRR (nucleotide-binding and leucine-rich repeat) proteins, which are subdivided into classes based on the N-terminal domain. In *Arabidopsis*, one class carries a TIR-domain (named due to similarity with *Drosophila* Toll and mammalian interleukin 1 receptors), the other class carries a coiled-coil (CC) domain. These two distinct N-terminal domains define specific signalling pathways that NB-LRRs initiate in response to effector recognition (Aarts et al., 1998). The importance of NB-LRRs in plant immunity is highlighted by the fact that the *Arabidopsis* Col-0 genome is estimated to encode $\approx$ 150 *NB-LRR* genes compared with mammalian genomes which encode $\approx$ 20 *NB-LRRs* (Meyers et al., 2003; Jones and Dangl, 2006; Jones et al., 2016).

Examples of well-studied *P. syringae* effectors include AvrRpt2 and AvrRpm1, which are recognised by the CC-NB-LRRs Resistant to *P. syringae* 2 (RPS2) and Resistant to *P. syringae* pv. *maculicola* 1 (RPM1) respectively. RPM1-interacting protein 4 (RIN4) plays a crucial role in perceiving both AvrRpt2 and AvrRpm1: RPM1 detects the phosphorylation of RIN4 by AvrRpm1 and RPS2 detects the cleavage of RIN4 by AvrRpt2, triggering ETI (Mackey et al., 2002, 2003). The cellular events following R-protein perception include ROS burst, calcium ion influx, activation of MAPK cascades, SA synthesis, transcriptional up-regulation of defence genes and the hypersensitive response (HR) (Dodds and Rathjen, 2010). HR is the rapid and localised programmed cell death (PCD) in infected cells in order to limit further pathogen invasion. The transcriptional changes induced by both defence strategies have been found to overlap extensively (Navarro et al., 2004), and there are many similarities between PTI and ETI. For example, both responses rely on activation of MAPK cascades for signalling; silencing of *Arabidopsis* MPK6 was found to compromise resistance against both the virulent *Pto* DC3000 and the avirulent *Pto* DC3000 expressing *avrRpt2* (Menke et al., 2004). Overall, it is thought that the differences between PTI and ETI do not result from alternate downstream signalling machinery; instead the two modes of defence differ in duration and amplitude of response (Tsuda and Katagiri, 2010).

Both PTI and ETI induce another facet of plant immunity known as systemic acquired resistance (SAR). This response induces a "priming" effect in distal plant foliage which is achieved by the systemic expression of *pathogenesis-related* (*PR*) genes, and requires the phytohormone salicylic acid (SA) as a signalling molecule (Fu and Dong, 2013).

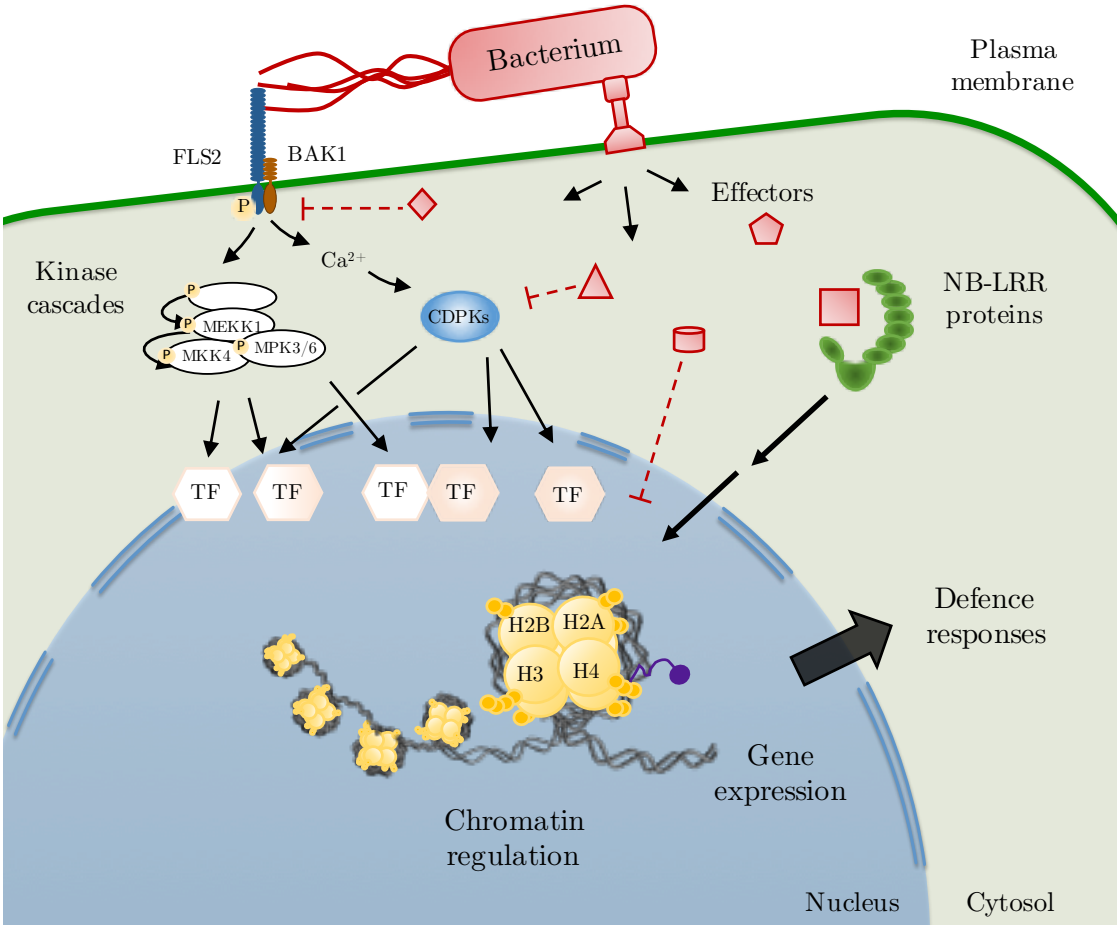


Figure 1.2: Signalling events and consequences following pathogen perception. Plants detect pathogen components, such as PAMPs and effectors, through PRRs and NB-LRRs causing the rapid activation of immune complexes. Signalling events, including CDPK and MAPK cascades, lead to activation of specific transcription factors which modulate differential gene expression to mount the appropriate level of defence response. The accessibility of DNA is dependent on chromatin structure, which is regulated by the action of ATP-dependent remodelling complexes and by histone covalent modifications. DNA is represented as grey lines, histones by yellow spheres and an example histone modification (such as acetylation) is depicted in purple.

1.3.4 Hormone signalling in defence

Plant hormones, such as brassinosteroids (BRs), ethylene (ET), abscisic acid (ABA) and jasmonic acid (JA), play a central role in defence against biotrophic and necrotrophic pathogens. Despite extensive research, the role and cross-talk of hormonal pathways remains a complex picture, although it is generally accepted that SA mediates defence responses against biotrophic pathogens and ET/JA mediates responses to necrotrophs (Glazebrook, 2005).

SA is a small phenolic compound whose synthesis is triggered upon pathogen perception (Durrant and Dong, 2004). In the absence of SA, the transcription co-factor NPR1 (Non-expressor of pathogenesis-related proteins 1) forms inactive oligomers in the cytoplasm. As SA concentrations increase, a redox change allows NPR1 to dissociate into a monomeric form and it translocates into the nucleus. NPR1 then binds TGA and WRKY family transcription factors to induce the expression of *PR* genes (Pieterse et al., 2009). NPR1 also plays a role in modulating cross-talk between SA and JA, as it inhibits the JA pathway in the cytosol (Spoel et al., 2003).

JA is a lipid-derived molecule with broad *in planta* roles: the hormone is known to regulate and signal in senescence, herbivore attack and necrotroph-responsive processes (Robert-Seilaniantz et al., 2011). In the absence of JA, jasmonate ZIM-domain-containing (JAZ) transcriptional repressors bind and repress the activity of target transcription factors, such as MYC2, MYC3 and MYC4 (Fernández-Calvo et al., 2011). The JA mechanism of action involves the degradation of the JAZ family through complex formation with COI1 (an E3 ligase), targeting the JAZ transcriptional repressor proteins for degradation (Sheard et al., 2010).

All hormone-mediated events are not independent from each other, and there is often cross-talk with other hormonal pathways, such as ethylene, cytokinin and brassinosteroid, to form complex hormonal interaction networks (Robert-Seilaniantz et al., 2011).

1.3.5 Defence and growth trade-off

Depending on external or internal cues, a plant will either prioritise growth or defence since activation of immunity requires substantial resources (Kempel et al., 2011; Meldau et al., 2012; Heidel et al., 2004). This trade-off has deep agricul-

tural implications, and due to this, the molecular mechanisms underlying this phenomenon are under intensive research. The cross talk of hormonal signalling pathways has been heavily implicated in fine-tuning of growth vs defence processes. For example, benzothiadiazole (BTH), a synthetic analogue of SA, is used to enhance disease resistance through activation of SAR in crops. Unwanted side-effects of BTH application to wheat were observed to include reduced growth and seed production (Heil et al., 2000). The major plant growth hormone, auxin, has been implicated in plant defence suppression since some pathogens can synthesise auxin directly or manipulate auxin signalling in the host (Chen et al., 2007; Glickmann et al., 1998). In response to this, plants have been shown to repress auxin signalling during defence: Navarro et al. (2006) found that after flg22 elicitation in *Arabidopsis*, *miRNA393* reduces the transcript and protein levels of auxin F-box receptors causing the stabilisation of AUX/IAA proteins which repress the expression of auxin-induced genes.

1.3.6 Transcriptional reprogramming during infection

One of the most fundamental outcomes of activation of defence processes is large-scale transcriptional reprogramming which allows effective defences to be mounted and coordinated. This response is known to be rapid: eliciting *Arabidopsis* with the PAMP flg22 causes the differential expression of approximately 10% of the genome within one hour after perception (Zipfel et al., 2004). Similarly, approximately 30% of the *Arabidopsis* genome becomes differentially expressed within 48 hours of infection with the necrotroph *Botrytis cinerea* (Windram et al., 2012). Recently, a high temporal-resolution transcriptomic experiment following infection with *Pto* DC3000 in *Arabidopsis* has allowed the dissection of transcriptional changes over the infection course (Lewis et al., 2015). Within the first two hours of infection, genes related to the characteristic ROS burst and responses to salicylic acid were up-regulated, whilst the expression of nuclear encoded chloroplast-targeted genes (NECGs) were suppressed (de Torres Zabala et al., 2015). Also, Lewis et al. (2015) demonstrated that transcripts associated with chromatin reorganisation were strongly suppressed in an effector-dependent manner, uncovering an early virulence strategy of host chromatin manipulation. The events leading to gene transcription are tightly regulated; complex networks consisting of positive and negative feedback loops allow signal amplification and tight regulation of the immune response.

1.4 Regulation of gene transcription in plants

Gene transcription is the process by which RNA polymerase produces mRNA from DNA. Subsequently, the mRNA molecule is used as a template to produce a corresponding protein in the process of translation (the central dogma of molecular biology (Crick, 1970)). Eukaryotic transcription is carried out by RNA polymerase II which complexes with a number of general transcription factors (GTF), such as transcription factor II D (TFIID), to initiate and carry out the process of transcription. TFIID, the first GTF to bind the gene to be transcribed, contains a TATA binding protein (TBP) subunit which recognises the TATA box within the upstream promoter region of a gene. After TFIID binds, transcription factor II A (TFIIA) and transcription factor II B (TFIIB) join and stabilise the polymerase complex. After the addition of several more GTFs to the complex, transcription proceeds upon ATP hydrolysis (Alberts, 2017).

The RNA polymerase complex is guided to target loci by transcription factors (TFs). TFs often contain DNA-binding domains which recognise specific promoter sequences (such as the WRKY family of TFs which recognise the W-box [(T)TGAC(C/T)] (Rushton et al., 2010)), and act to modulate the activation or repression of transcription at distinct gene loci (Luscombe et al., 2000). The mechanisms of transcriptional activation and repression are varied and complex. One example of a repressive mechanism is the promotion of histone deacetylation causing coding regions and promoters to become inaccessible to RNA polymerase machinery (Grunstein, 1997). The positive or negative regulation of gene transcription is tightly linked with chromatin structure which is controlled by chromatin modifying proteins.

1.4.1 Chromatin remodelling

The DNA macromolecule is tightly packed inside the nucleus to a millionth of its length. This incredible feat is achieved through the wrapping of DNA around small (11-21 kDa), positively charged proteins known as histones which are highly conserved in all eukaryotes. The basic unit of chromatin is a histone octamer: two dimers of H2A-H2B (histone 2A and histone 2B) and one tetramer of H3-H4 (histone 3 and histone 4) with approximately 146 bases of DNA wrapped around each octamer completes the nucleosome structure. Linker DNA is found in the space between nucleosomes where histone H1 associates to strengthen higher-order packing

(Kouzarides, 2007; Clark and Kimura, 1990). Chromatin is largely found in two states: transcriptionally silent and condensed heterochromatin, and transcriptionally active, "loose" euchromatin.

Crystallographic studies have shown that the N-terminus of histone tails (30% of the protein mass) protrude outwards from the nucleosome (Luger et al., 1997). The tails, which are lysine-rich, can undergo various types of post-translational modifications (PTM) such as methylation, acetylation and phosphorylation (although mass spectrometry analysis has discovered more than 200 other novel modifications (Macek et al., 2006)). The modifications serve a multitude of roles, but on a most basic level they can strongly and rapidly affect chromatin structure by directly changing DNA/histone interactions within the nucleosome or indirectly affect chromatin structure by recruiting histone-modifying complexes (Wegel and Shaw, 2005; Kouzarides, 2007). Chromatin structure can also be affected by ATP-dependent remodelling complexes, of which there are more than 40 putative genes encoded in the *Arabidopsis* genome (Clapier and Cairns, 2009) (Fig. 1.3). These complexes have reader domains that interact with nucleosomes in order to re-structure chromatin compaction through moving, ejecting or restructuring nucleosomes (Clapier and Cairns, 2009). Direct DNA (cytosine) methylation is also known to affect chromatin compaction, and is highly correlated with transcriptionally silent areas of the genome known as heterochromatin (Alvarez et al., 2010). Small non-coding RNAs are known to direct and establish heterochromatin formation at specific gene loci (Chan et al., 2004). All cellular processes that involve DNA (DNA-repair, replication and transcription) are influenced by local and higher order chromatin structure.

1.4.2 Histone modifications

The deposition of histone modifications is catalysed by a variety of histone-modifying enzymes which are recruited to specific genomic regions by transcription factors (Kouzarides, 2007). Different histone modifications are associated with different levels of gene expression. For example, acetylation of histones is associated with "loosening" the interactions between histone proteins and DNA, allowing these regions to become more accessible to DNA regulatory machinery. This occurs through alterations to local electrostatic charges: the positive charge of histone tail lysines are neutralised through the modification of an acetyl group. Since DNA is nega-

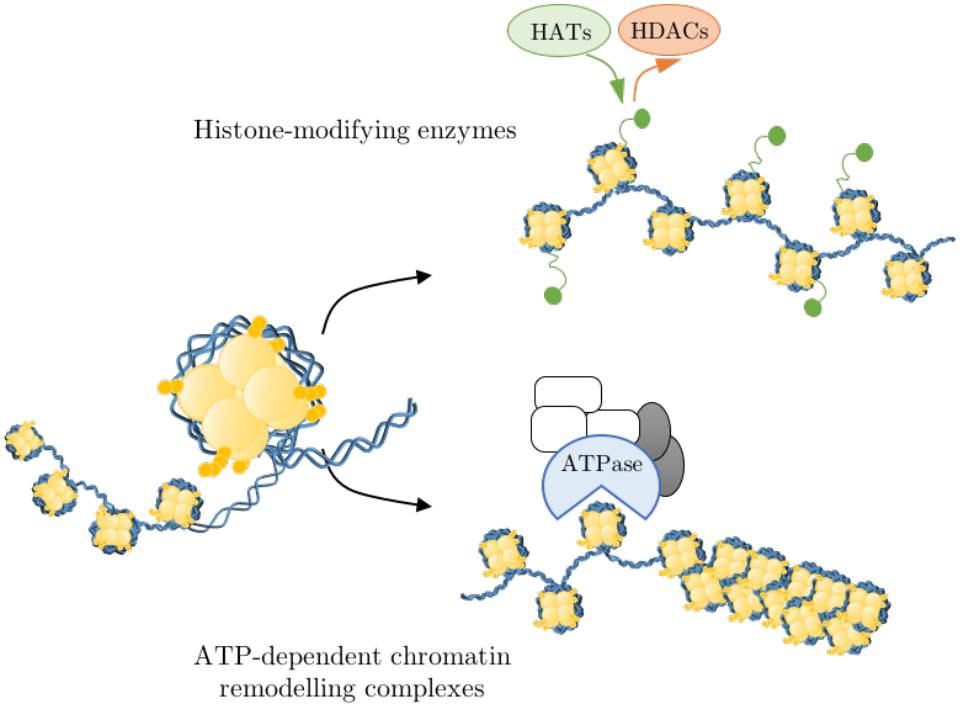


Figure 1.3: Histone modifying enzymes and chromatin remodelling complexes determine chromatin structure. Histone modifying enzymes (top), such as histone acetyltransferases (HATs), affect chromatin structure by deposition of acetyl groups onto lysines of histone tails (represented in green). This modification is removed by the antagonistic action of histone deacetylases (HDACs). Other histone tail modification include phosphorylation, methylation and ubiquitination. ATP-dependent chromatin remodellers (bottom) hydrolyse ATP via ATPase domains to slide or replace nucleosomes. Histone modifying enzymes and complexes do not act in an exclusive manner, but rather in concert to determine chromatin structure.

tively charged, the addition of an acetyl group weakens DNA-histone interactions which can affect the local and higher-order chromatin structure. Acetyl groups are also known to act as docking site for histone-interacting proteins. Some histone acetyltransferases contain bromodomains that bind acetylated lysines, propagating the acetylation modification to other local histone tail lysines.

The placement of one histone mark is not enough to define a transcriptionally active or silent region. In fact, the level of transcription can depend on the number and combination of marks in that genomic region. For example, it was found that phosphorylation of serine-10 of H3 (H3S10) enhances the acetylation of H3K13 in *Saccharomyces cerevisiae* (Trievel et al., 1999). For plants in general, regions of

transcriptional silence are associated with methylated H3 lysine 27 (H3K27), hypoacetylation of H3 and H4, and hypermethylated DNA. Actively transcribed regions show H3 and H4 hyperacetylation, heightened trimethylation of lysine 4 of H3 (H3K4me3) and DNA hypomethylation (Roudier et al., 2011; Zhang et al., 2009)

1.5 Histone acetylation

1.5.1 *Arabidopsis* histone acetyltransferases

Histone lysine acetylation status is controlled antagonistically by histone acetyltransferases (HATs) and histone deacetylases (HDACs). HATs use the co-factor acetyl-CoA to catalyse the transfer of the acetyl moiety of acetyl-CoA to the ε-amino group; forming CoASH (coenzyme A with a thiol group) as a by-product. With the addition of a water molecule, HDACs hydrolyse the acetyl group, leading to its removal (Berger, 2007) (Fig. 1.4). The *Arabidopsis* genome contains 12 HATs and 16 HDACs (Pandey et al., 2002). They are involved in a variety of biological processes including development, response to abiotic stress (Earley et al., 2007), flowering time (Xiao et al., 2013), responses to light (Bertrand et al., 2004) and JA/ET and SA signalling pathways (Servet et al., 2010).

Arabidopsis HDACs can be subdivided into three families: (1) the RPD3/HDA1 super-family, (2) the Silent Information Regulator 2 (SIR2) family and (3) the HD2 family. HDACs are often associated with hypoacetylated genomic regions, and transcriptionally silent genes are often linked with a multiprotein complex containing a HDAC (Liu et al., 2014). In the context of plant immunity, the role of HDA19 (which belongs to the RPD3/HDA1 super-family) has been well described. Zhou et al. (2005) found that *HDA19* knock-down lines had reduced expression of ET/JA pathway genes and increased susceptibility to necrotrophic pathogens. HDA19 was subsequently also shown to positively-regulate SA-mediated basal defences: Kim et al. (2008) demonstrated that expression of *HDA19* was induced upon *Pto* DC3000 infection, whereupon HDA19 indirectly caused the de-repression of defence gene expression. *hda19* mutant lines were also found to be more susceptible to *Pto* DC3000, indicating that this HDAC positively contributes to defences. HDT701, a rice histone deacetylase from the HD2 family, was found to negatively regulate innate immune responses by deacetylating the promoters of defence-related

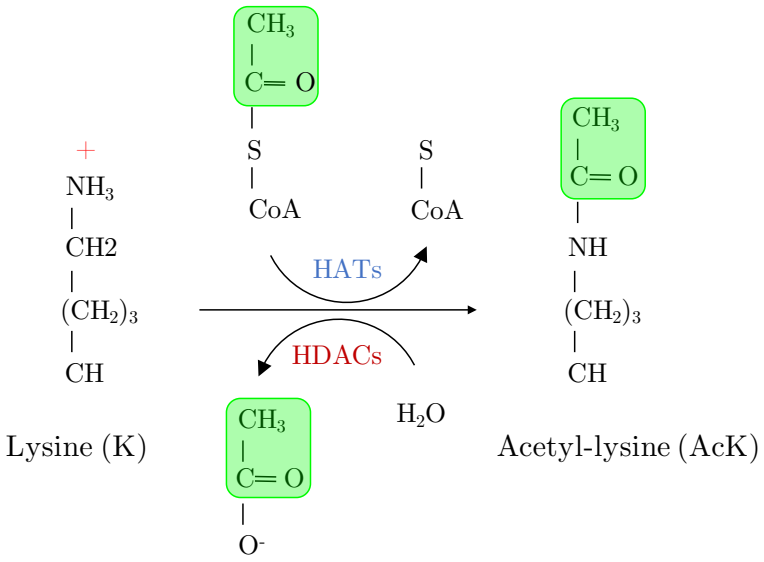


Figure 1.4: Acetylation and de-acetylation of lysine residues. The transfer of an acetyl group (green box) to a lysine residue is catalysed by histone acetyltransferases, forming coenzyme A with a thiol group (CoASH) as a by-product. Histone deacetylases hydrolyse the acetyl group leading to its removal. The mechanism for catalysis, and the residues utilised during the reaction, differ between the HAT subfamilies. Modified from Kim and Yang (2011).

genes. Consequences of HDT701 silencing included an enhanced ROS burst after PAMP elicitation and resistance to the necrotrophic pathogen *Magnaporthe oryzae* (Ding et al., 2012). A more recent example includes Latrasse et al. (2017), who were able to show the direct interaction of MPK3 and the *Arabidopsis* HDAC HD2B, regulating the localisation and function of the histone deacetylase. It is therefore clear that regulators of histone acetylation levels are major contributors in the reprogramming of defence gene expression.

The 12 *Arabidopsis* HATs are subdivided into four families based on comparative sequence analyses: GNAT (HAG1, HAG2, HAG3), p300/CBP (HAC1, HAC2, HAC4, HAC5, HAC12), MYST (HAM1 and HAM2) and $TAF_{II}250$ (HAF1 and HAF2) (Pandey et al., 2002). The HATs exist as large, multi-domain proteins and the mechanism of catalysis differs between each HAT sub-family. The domain predictions for each HAT protein, along with the target specificity (as determined by *in vitro* (Earley et al., 2007) and *in vivo* assays) are presented in Figure 1.5. Here it can be seen that all HATs contain a family-specific HAT domain (highlighted in green),

along with extra domains that determine binding partner specificity for that enzyme.

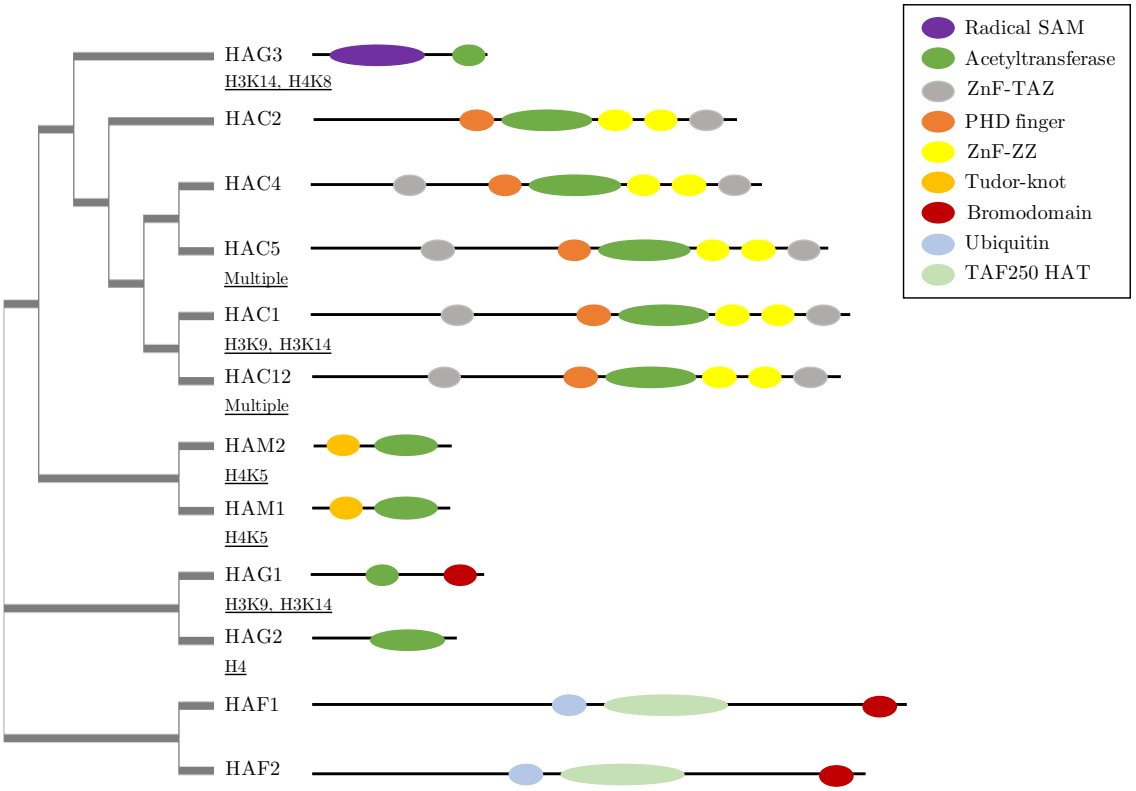


Figure 1.5: Phylogenetic tree and domain architecture of the *Arabidopsis* HATs. The phylogenetic tree was constructed after multiple sequence alignment of all HAT amino acid sequences using Clustal Omega (Sievers et al., 2011) with neighbour joining clustering. For domain architecture, information was collected from UniProt (Consortium, 2016). The length of the horizontal bars depicting each protein is proportional to the protein sequence length. For histone target specificity (underlined), information was collated from Earley et al. (2007); Singh et al. (2014); Bertrand et al. (2004); Xiao et al. (2013).

GNAT family: HAG1, HAG2 and HAG3

HAG1, also known as GCN5, is perhaps the most well-described *Arabidopsis* HAT. It has been shown to play a key role as a transcriptional co-activator in many processes including cell differentiation promotion, leaf development and root meristem differentiation (Servet et al., 2010). In more recent publications, Poulios and Vlachonasios (2015) uncovered a role for HAG1 in inhibition of ethylene responses in *Arabidopsis*, and Wang et al. (2016) demonstrated that HAG1 modulated fatty acid

biosynthesis by affecting acetylation levels of *FAD3* (*Fatty Acid Desaturase 3*). *Arabidopsis* lines with T-DNA insertions in *HAG1* show a variety of pleiotropic defects including dwarfism, aberrant meristem function and floral defects affecting fertility (Vlachonasios et al., 2003; Benhamed et al., 2006; Kornet and Scheres, 2009). The importance of this HAT became clear when Benhamed et al. (2008) performed ChIP coupled with promoter ChIP to determine promoter-occupancy of HAG1. It was found that approximately 40% of *Arabidopsis* promoters were associated with HAG1. It is also interesting to note that HAG1 is known to acetylate ADA2, a member of the SAGA complex of which HAG1 is the catalytic subunit (Mao et al., 2006). This may indicate that HAG1 self-regulates its own activity through the acetylation of ADA2. Also, it was recently found that PsAvh23, an effector deployed by the soybean pathogen *Phytophthora sojae*, binds the ADA2 subunit of the SAGA HAT complex. This was shown to interfere with the necessary association with the catalytic subunit GCN5, leading to a reduction of H3K9 acetylation levels at defence-gene loci and increased host colonisation by *Phytophthora sojae* (Kong et al., 2017).

Another GNAT HAT, HAG3, was found to have a role in UV-B induced DNA damage repair and signalling, since *hag3* RNAi lines show higher levels of UV-B-absorbing compounds and less UV-B-induced DNA damage than the wild-type control (Fina and Casati, 2015). Interestingly, a role was also found for HAG3 in defences: it was found to be a positive regulator of plant immunity by accelerating the expression of defence-related genes such as *PR1* which was dependent upon its catalytic acetyltransferase activity (DeFraia et al., 2013).

p300/CBP family: HAC1, HAC2, HAC4, HAC5 and HAC12

The p300/CBP family of HATs contain five family members with broad-specificity H3 lysine acetylation activity (Earley et al., 2007). Li et al. (2014a) analysed the morphological and developmental phenotypes of *hac* mutants and found that HACs are involved in pleitropic developmental processes including plant size, leaf morphology and flowering time. By creating single, double and triple *hac* mutants, they found that *HAC1* played the most dominant role with the synergistic assistance from *HAC5*, *HAC4* and *HAC12* (Li et al., 2014a). Interestingly, another role has been described for HAC1: repeatedly challenging a plant with flg22 normally induces a priming response, causing the plant to become more resistant to infection. After repeated challenging with flg22, *hac1* lines failed to establish a primed state,

open chromatin states and showed no resistance to bacterial infection (Singh et al., 2014).

$TAF_{II}250$ family: HAF1 and HAF2

HAF2 was found to have a fundamental role in integrating light signals and histone acetylation: ChIP assays were performed on mutated *HAF2* lines and a reduction in the acetylation levels of H3 could be seen in light-responsive promoters (Bertrand et al., 2004). Also, when screening *HAT* mutant lines, Chen et al. (2016) discovered a role for HAF2 in root epidermal patterning; altered epidermal phenotypes were seen for *haf2* lines. To date, there have been no publications linking the members of the *Arabidopsis* HAF family to plant defence processes.

MYST family: HAM1 and HAM2

Several publications have described a role for the two *Arabidopsis* MYSTs in distinct processes such as gametophyte development (Latrasse et al., 2008), UV damage response (Campi et al., 2012) and drought responses in rice (Fang et al., 2014). To this date, there have been no publications describing a link between HAMs and plant defence. A role for HAMs in gametophyte development was uncovered by Latrasse et al. (2008), who found that *ham1/ham2* null mutants were not viable. Following generation of a *ham* sesquimutant (*HAM1/ham1; ham2/ham2* and *ham1/ham1; HAM2/ham2*), they found these lines presented defects in silique length, seed number and formation of both male and female gametophytes. Following this, Campi et al. (2012) analysed the responses to UV-B treatment in *hat* mutants. It was found that single *ham1* and *ham2* mutants both had increased DNA damage after UV-B compared to wild-type. The accumulation of cyclobutane pyrimidine dimers was higher in *ham1* than *ham2*, suggesting a more dominant role for HAM1 in DNA damage repair after UV-B. Fang et al. (2014) analysed the responses of rice HATs under drought stress conditions, and found that expression of the rice MYST HAT, OsHAM701, was induced under drought stress. They also found that levels of H4K5 increased under drought stress (H4K5 was found to be targeted for acetylation by *Arabidopsis* MYSTs by Earley et al. (2007)).

The MYST family of HATs are evolutionary conserved throughout eukaryotes and, in yeast and mammalian systems, have been shown to regulate diverse cellular processes such as DNA repair, stem cell homeostasis and cell-cycle regulation through

acetylation of histone 4 (H4) lysine residues (Yuan et al., 2012). Although acetylation of H3 tail residues is associated with actively transcribed regions, it has been shown that acetylation of H4 lysine residues has distinct functions that do not correlate with induction of gene expression. For example, recently Kaimori et al. (2016) studied the function of H4 lysine 20 acetylation (H4K20ac). Following ChIP-seq analysis in human cells, it was found that H4K20ac was enriched in the transcription start sites (TSSs) of minimally expressed genes. Motif over-representation analysis of H4K20ac-enriched sequences revealed that transcriptional activators were excluded from these area; instead the transcription repressor NRSF/ REST (neuron-restrictive silencer factor/repressor element 1-silencing transcription) co-localized with H4K20ac. In other study, the acetyltransferase activity of Chameau (Chm) (a member of the *Drosophila* MYST HAT family) was found to be required for the maintaining of Hox gene silencing (Grienenberger et al., 2002). Indeed, KAT5 (a MYST HAT in humans) was found to repress basal expression when fused to a reporter gene and was found to interact with histone deacetylase 7 (HDAC7), which potentiated the transcriptional repression of STAT3, a transcriptional activator (Xiao et al., 2003). It can therefore be seen that transcription at a particular genomic loci is not determined by overall acetylation level, but rather the specific combination of histone modifications at the location.

1.6 Aims of this study

As described above, many publications have described a role for HATs and HDACs in developmental, morphological and other processes. Also, several reports have uncovered a link between HAT/HDAC activity and plant immunity processes. Based on these publications, it is clear that the major transcriptional changes that occur during pathogen infection require the activity of regulators of histone acetylation.

In this work, the role of *Arabidopsis* HATs in the different phases of plant immunity were investigated using a reverse genetics approach. The results in Chapter 3 identify HAM2, a MYST family member, to be a negative regulator of basal immunity. The *ham2* phenotypes observed present HAM2 as a novel target in an agricultural context. Therefore, in Chapter 4, a homology modelling approach was taken to predict the protein structure of *Arabidopsis* HAM2 (as well as several HAM2 homologues in important crop species). An *in silico* docking screen was then performed to identify chemical inhibitors targeting HAM2. In Chapter 5, *Arabidopsis hat* mu-

tants were screened for misregulation of ETI. This resulted in the finding that HAG1 is a key positive regulator of ETI processes, as major outputs of ETI were severely affected in the *hag1* mutant.

Chapter 2

Material and Methods

2.1 Plants used in this study

2.1.1 Plant growth conditions

Arabidopsis thaliana seeds were sown on Arabidopsis mix (F2 compost, Intercept, grit) and stratified in darkness for 2-3 days at 4°C. Seedlings were germinated in a controlled environment chamber at 20°C, 60% humidity in short-day 10/14 hour (day/night) cycles with a light fluence rate of 100 μmol/m^2/s. When 12-14 days old, seedlings were transferred to individual pots and maintained with the same growth conditions as above. When seeds were required, plants were transferred to a growth chamber with a long-day 16/8 hour (day/night) cycles to promote flowering.

For *in vitro* experiments, seeds were surface sterilised before plating. In a sealed desiccator, seeds were exposed to chlorine gas (3 mL of 37% HCl added to 100 mL of 10% sodium hypochlorite to create the gas) for 4 hours. Seeds were transferred to sterilised 1/2 MS plates containing Murashige & Skoog medium (2.15 g/L, Duchefa Biochemie), 1% sucrose (Sigma) and 0.5% Phytagel (Sigma) (pH 5.8) (Phytagel added when solid media required). All plating procedures were carried out aseptically in a Class II sterile cabinet. Seedlings were stratified in darkness for 2-3 days at 4°C and then germinated in a controlled environment chamber with the same growth conditions as stated above.

2.1.2 *Arabidopsis* mutant lines

All *Arabidopsis thaliana* T-DNA insertion lines used in this study are of a Columbia-0 (Col-0) background and are displayed in Table 2.1. SALK lines were obtained from the Nottingham Arabidopsis Stock Centre (NASC, http://arabidopsis.org.uk/). Homozygote T-DNA insertion lines were identified by genotyping PCR using primers that anneal to the gene-specific sequence and the left border of the T-DNA insert (Alonso et al., 2003). All primers used in this study can be found in Appendix A.

Table 2.1: *Arabidopsis thaliana* mutant lines used in this study.

***Arabidopsis* line**	**AGI number**	**T-DNA insertion line**	**Used in**
hag1-6	At3g54610	SALK_150784	Chap. 3
hag2-99	At5g56740	SALK_051832	Chap. 3
haf2-29	At3g19040	SALK_110029	Chap. 3
ham1-50	At5g64610	SALK_103726	Chap. 3
ham1-96	At5g64610	SALK_027726	Chap. 3
ham2-71	At5g09740	SALK_012086	Chap. 3, 4, 5
ham2-75	At5g09740	SALK_106046	Chap. 3, 4, 5
hac2-84	At1g67220	SALK_049434	Chap. 3
hac4-15	At1g55970	SALK_051750	Chap. 3
hac4-21	At1g55970	SALK_045791	Chap. 3
hac5-78	At3g12980	SALK_024278	Chap. 3
hac12-04	At1g16710	SALK_071102	Chap. 3

2.1.3 Generation of *Arabidopsis* crosses

Unopened flower buds with immature stamen and mature stigma were identified for crossing. These flowers were emasculated using sterile fine tweezers and pollen from donor stamen was brushed against the recipient stigma. Siliques from the cross were allowed to dry and the F_1 progeny were harvested. F_1 seeds were sown and allowed to self-pollinate, after which F_2 progeny were harvested.

2.1.4 Stable transformation of *Arabidopsis*

Transgenic *Arabidopsis* plants were produced using the floral dip method (Clough and Bent, 1998). In brief, flowering *Arabidopsis* plants were dipped into a liquid culture of *Agrobacterium tumefaciens* carrying the plasmid to be transformed (*A. tumefaciens* transformation detailed in Section 2.5.5). T_1 seeds were harvested and sown on rockwool soaked in 1/2 MS and 5 μg/mL Basta. Basta-resistant seedlings were transplanted and T_2 seed generated. T_2 plants were confirmed to carry the transgene of interest by PCR, and protein expression was checked by western blot analysis.

2.2 Plant pathogen material

2.2.1 *P. syringae* material and growth

Prior to infection assays, bacteria were streaked from stock (maintained at -80°C in 20% glycerol) onto selective media plates and grown at 28°C for two days. Single colonies were used to inoculate 10 mL liquid cultures which were grown overnight. *P. syringae* strains were grown in sterile King's B medium (20 g/L Bacto Peptone, 8.60 mM K_2HPO_4, 163 mM glycerol and 0.5% agar (pH 7) (agar added when solid media was required). All isolates used in this study are from *Pseudomonas syringae* pv. *tomato* strain DC3000 and are listed in Table 2.2.

2.2.2 *B. cinerea* material and growth

B. cinerea var. pepper (Denby et al., 2004) was cultured on sterilised apricot halves at 25°C. Spores were collected 14 days post inoculation by scraping fungal material into 3 mL sterile H_2O in a Class II sterile cabinet. The spore solution was filtered through glass wool to remove debris, spores were counted using a haemocytometer and the solution concentration was adjusted to 2×10^5 spores/mL. The solution was then diluted 1:1 to a final concentration of 1×10^5 spores/mL with sterile grape juice.

Table 2.2: *Pseudomonas syringae* pv. *tomato* strains used in this study.

Strain	Open Reading Frame	Vector	Selection
DC3000 (wild-type)	n/a	n/a	Rifampicin 100 μg/mL
DC3000	Empty Vector	pVSP61	Rifampicin 100 μg/mL Kanamycin 25 μg/mL
DC3000	AvrRpt2	pVSP61	Rifampicin 100 μg/mL Kanamycin 25 μg/mL
DC3000	AvrRpm1	pVSP61	Rifampicin 100 μg/mL Kanamycin 25 μg/mL
DC3000	AvrPphB	pVSP61	Rifampicin 100 μg/mL Kanamycin 25 μg/mL
DC3000	AvrRps4	pVSP61	Rifampicin 100 μg/mL Kanamycin 25 μg/mL

2.3 Plant pathology assays

2.3.1 Flg22 peptide elicitation

Seedling growth inhibition

For each treatment, 12 seedlings were grown in sterile conditions as described in Section 2.1.1. Seedlings were initially germinated on solid 1 % sucrose 1/2 MS medium and allowed to grow for 6 days. After this, the seedlings were then transferred to liquid 1 % sucrose 1/2 MS medium with 0 nM, 1 nM, 10 nM or 100 nM flg22 (Peptron, South Korea). Fresh weight of each seedling was measured one week later.

2.3.2 *P. syringae* infection

Infiltration

P. syringae overnight liquid cultures (from a single bacterial colony) were harvested by centrifugation, washed and re-suspended in sterile 10 mM $MgCl_2$. Cultures were diluted by serial dilution to 2×10^5 cfu/mL (optical density at 600 nm (OD_{600})= 0.001) in 10 mM $MgCl_2$. Cell density measurements were taken using a Biochrom WPA CO8000 cell density meter (Biochrom Ltd., UK). Three leaves of 4-5 week old *Arabidopsis* plants (leaves 7, 8 and 9) were infiltrated using a needleless 1 mL syringe. Bacterial population sizes in infected leaves were quantified three days post

inoculation, unless stated otherwise. For ion leakage assays, cultures at 2×10^7 cfu / mL (OD_{600}= 0.1) were used for infiltration.

Spraying

P. syringae overnight liquid cultures (from a single bacterial colony) were harvested by centrifugation, washed and re-suspended in sterile 10 mM $MgCl_2$. Cultures were diluted to 2×10^7 cfu / mL (OD_{600}= 0.1) in 10 mM $MgCl_2$ and Silwet-L77 (Lehle Seeds) was added to the suspension to 0.04%. Four-to-five week old plants were sprayed until all leaves were visibly soaked (Sparmax TC-620X spray paintbrush (The AirbrushCompany)). Six plants per genotype were infected per experiment. Plant trays were covered with plastic lids to maintain approximately 90% humidity for 24 hours. Lids were then removed and plants were kept at standard growth conditions detailed in Section 2.1.1. Bacterial population sizes in infected leaves were quantified three days post inoculation, unless stated otherwise.

2.3.3 Quantification of bacterial growth in infected leaves

Twelve leaf punctures per genotype were taken using a No.4 cork borer from infected leaves. Two leaf punctures from different plants were pooled into one 2 mL tube containing 200 μL sterile 10 mM $MgCl_2$ and 2 metallic beads. Leaf tissue was ground through two pulses of 28 Hz for 30 seconds in a tissue lyser (MM300, Retsch). Eight-hundred μL 10 mM $MgCl_2$ was added to each suspension to make the total volume 1 mL. Serial dilutions of each suspension were created and plated onto KB plates containing the appropriate antibiotics. All plating procedures were carried out aseptically in a Class II sterile cabinet to avoid contamination. The number of bacterial colonies were counted 48 hours later.

2.3.4 Ion leakage measurement

Induction of the hypersensitive response (HR) leads to localised programmed cell death resulting in the release of ions from dying cells (Goodman et al., 1994). The amount of electrolytes released from leaves can therefore be used to quantify HR cell death. As described in Section 2.3.2, for this assay leaves number 7, 8 or 9 were infiltrated with *P. syringae* strains at OD_{600}= 0.1. Six plants per line were infected, with 2 leaves per plant infiltrated. Within 30 mins of infiltration, twelve leaf punctures per genotype were taken with a No.4 cork borer. These were pooled in 40 mL sterile H_2O and gently shaken for 1 hour. 2 leaf discs per well were

then transferred to 24-well plates containing 2 mL sterile H_2O. Conductivity of well contents was measured over a time-course using a conductivity meter (B-173 Twin Cond, Horiba).

2.3.5 *B. cinerea* infection

Leaves 7, 8 and 9 of 4-5 week old plants were detached and placed on 0.8% Phytoagar (Duchefa Biochemie) trays. A minimum of 42 leaves per plant genotype were used, and each genotype was distributed across several trays to reduce variability caused by tray differences. A 5 μL droplet of 1×10^5 spores/mL solution was placed in the middle of each leaf. Trays were covered with lids and sealed to maintain approximately 90% humidity. Leaves were photographed every 24 hours post inoculation for 3 days and lesion sizes were measured and analysed using ImageJ (Schneider et al., 2012).

2.4 Plant development assays

2.4.1 Leaf surface area and fresh weight measurement

Arabidopsis plants were grown as detailed in Section 2.1.1. Photographs were taken 4 to 5 weeks post-germination (alongside a ruler for scale). Leaf surface area was measured and analysed using ImageJ (Schneider et al., 2012). A minimum of 12 plants were measured for each experimental condition. For fresh weight measurements, the rosettes of 5 week-old *Arabidopsis* plants were weighed on a balance.

Arabidopsis seedlings were grown *in vitro* as detailed in Section 2.1.1. The weight of the whole seedling (root and shoot) was measured 2 weeks post-germination. A minimum of 12 plants were measured for each experimental condition.

2.4.2 Primary root measurement

Arabidopsis seedlings were grown *in vitro* on solid 1% sucrose 1/2 MS plates as detailed in Section 2.1.1. A minimum of 12 seedlings were measured for each experimental condition. At 15 days old, plates were scanned, alongside a ruler for scale, using a HP PSC 2500 scanner at 1200dpi and saved as ".tif" files. Primary root length was measured using ImageJ (Schneider et al., 2012) using the freehand measure tool.

2.4.3 Root apical meristem measurement

Arabidopsis seedlings were grown *in vitro* on 1% sucrose 1/2 MS as detailed in Section 2.1.1. Five-seven days after germination, 1 cm root tips were cut and immersed in a solution of 20 mg/mL propidium iodide (Sigma) followed by 10 minutes incubation in the dark. Root tips were mounted on a microscope slide with water and imaged using a Zeiss Laser Scanning Microscope 710 (Carl Zeiss Ltd) within 30 minutes of root cutting. At least 12 root tips were imaged per treatment. Root apical meristem size was expressed as the number of cells extending from the quiescent centre to the first elongated cortex cell.

2.5 Molecular biology methods

2.5.1 DNA methods

Arabidopsis DNA extraction for genotyping

Approximately 4mm diameter of leaf tissue was homogenised in 100 μL 5% Chelex 100 resin (Bio-Rad; (HwangBo et al., 2010)), vortexed then incubated at 99°C for 5 min. Samples were vortexed and spun down to pellet debris. One μL was used for genotyping PCR.

E. coli and *P. syringae* plasmid extraction

Overnight liquid cultures were harvested by centrifugation. Plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions.

2.5.2 Polymerase chain reaction (PCR)

Genotyping PCR

A 20 μL genotyping PCR reaction mix consisted of 1 μL gDNA, 1 μL 10 μM primer F, 1 μL 10 μM primer R, 0.5 μL 10 mM dNTPs, 0.2 μL KAPA Taq DNA polymerase (KAPA Biosystems), 4 μL 5 x KAPA Taq buffer and 12.3 μL H_2O. Genotyping PCR programs were performed using a thermal cycler (PTC-225 Peltier Thermal Cycler, MJ Research) with the conditions detailed in Table 2.3. Primers used to genotype T-DNA insertion lines can be found in Appendix A.

Table 2.3: Thermal cycling conditions used for genotyping PCR reactions.

Step	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	30
Annealing	55	20 sec	
Elongation	72	30 sec	
Final elongation	72	4 min	1
Cooling	4	Hold	

High-Fidelity PCR

This method of PCR was used for cloning when sequence accuracy was required. PCR reaction mixes contained 5 μL 5 x Q5 reaction buffer, 0.5 μL 10mM dNTPs, 1.25 μL 10 μM primer F, 1.25 μL 10 μM primer R, 1 μL template DNA, 0.25 μL Q5 High-Fidelity DNA polymerase (New England Biolabs) and nuclease-free H_2O to 25 μL.

Colony PCR

Colony PCR was used to confirm insert size when cloning. Bacterial colonies were picked and diluted in 50 μL sterile H_2O, then incubated at 95°C for 10 mins. One μL was added to each PCR reaction which is detailed in Section 2.5.2.

2.5.3 Gel electrophoresis

PCR products were separated by electrophoresis on agarose gel consisting of 1% agarose (Sigma), 1 x TAE buffer (40 mM Tris base, 1 mM EDTA, 20 mM acetic acid, pH 8.0) and 1 x GelRed (Biotium). Gels were visualised on a UV transilluminator (Gel Doc 1000, Bio-Rad).

2.5.4 DNA extraction from agarose gel

Amplicons were visualised on a UV transilluminator (TM40, UVP) and the desired fragment was cut carefully using a razor blade. The fragment was then purified using

QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions.

2.5.5 Gateway cloning

Fragment amplification

To generate inserts compatible for Gateway recombination reactions, a two-step PCR was used to incorporate attB sites into the DNA fragments. Gene-specific primers were designed and contained the following mini attB1/attB2 sites:

Forward: 5'- AAAAAGCAGGCTATG -3'
Reverse: 3'- AGAAAGCTGGGTC -5'

High-fidelity PCR was performed (as described above) using the thermal cycling conditions described in Table 2.4. After checking amplicon size on a 1% agarose gel, PCR products were cleaned using QIAquick PCR Purification Kit (Qiagen) following manufacturer's instructions. If more than one amplicon was produced after PCR, gel extraction was performed to isolate the desired fragment as detailed in Section 2.5.4.

Zero point one ng of this purified PCR product was used as a template for a second PCR to incorporate full attB sites at the ends of fragments. The same thermal cycling conditions were used as described in Table 2.4 using the full length attB1/attB2 primers:

attB1: 5'- GGGGACAAGTTTGTACAAAAAAGCAGGCT -3'
attB2: 3'- GGGGACCACTTTGTACAAGAAAGCTGGGT -5'

BP and LR reactions

Fragments containing the full attB sites were cloned into the entry vector pDONR/Zeo with Gateway BP Clonase II Enzyme mix (Invitrogen) following the manufacturer's protocol.

After colony selection and plasmid isolation, pDONR/Zeo containing the desired insert was subjected to a second recombination reaction using Gateway LR Clonase II Enzyme mix (Invitrogen) into the destination vector pBAV154 (which contains a

dexamethasone-inducible promoter and a C-terminal HA tag (Vinatzer et al., 2006). The manufacturer's protocol was followed for this reaction.

Table 2.4: Thermal cycling conditions used for attB PCR reactions.

Step	**Temperature (°C)**	**Time**	**Cycles**
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	5
Annealing	55	20 sec	
Elongation	72	50 sec/kb	
Denaturation	98	10 sec	25
Annealing and elongation	72	3 min 20 sec	
Final elongation	72	2 min	1
Cooling	4	Hold	

Electroporation of *E. coli*

Electrocompetent TOP10 *E. coli* cells (Invitrogen) were thawed on ice for 5 mins. Two point five μL plasmid was added to 50 μL of thawed cells and transferred to an electroporation cuvette. Cells were electroporated using the following settings: 1800V, capacity 25 μF and 200 Ω resistance in an MicroPulser Electroporator (Bio-Rad). Cells were re-suspended in 500 μL of LB liquid medium and incubated at 37°C in a shaker for 1 hour. The cell suspension was then plated onto agar plates containing appropriate antibiotics.

A. tumefaciens heat-shock transformation

One μg of plasmid was added to *A. tumefaciens* GV3101 competent cells and incubated on ice for 30 minutes. Cells were heat-shocked in liquid nitrogen for 5 min, followed by incubation at 37°C for 5 min and put on ice for another 5 minutes. One mL LB was added followed by shaking incubation for 2.5 hours at 28°C at 210 rpm. Colonies were selected for on LB agar plates containing appropriate antibiotics.

2.5.6 DNA sequencing

DNA quality and concentration was determined with a NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Fisher Scientific). DNA sequencing was carried out by GATC Biotech (Germany) and sequencing mixes contained 80-100 ng/μL DNA and 5 pmol/μL primer. Primers used for sequencing can be found in Appendix A.

2.5.7 RNA methods

Total RNA isolation

After storing at -80°C, samples were homogenised by pestle and mortar to a fine powder in liquid nitrogen. The tissue was immediately transferred to a 2 mL tube and kept in liquid nitrogen.

In a fume hood, 1 mL of Trizol reagent (Sigma) was added to each tube of plant material, vortexed and left at room temperature for 5 minutes. Two hundred μL chloroform was added, samples were inverted and left at room temperature for a further 5 minutes. The sample solution was centrifuged at 12000 x g for 20 min and the upper phase was transferred to a clean tube containing an equal volume of cold isopropanol. Samples were incubated overnight at -20°C and then centrifuged at 12,000 x g for 20 min. RNA pellets were washed twice with RNase-free 70% ethanol, air-dried and resuspended in RNase-free water. The quality and yield of the RNA was assessed by running on a 1% agarose gel and by measuring with a NanoDrop ND-1000.

cDNA synthesis

RNA was reverse-transcribed with the SuperScript II Reverse Transcriptase (Thermo Fisher Scientific). To make a 20 μL reaction, 2 μg RNA, 0.5 μL oligodT and 1 μL 10 mM dNTPs were incubated at 65°C for 5 min. Then 4 μL 5 x First-Strand buffer and 2 μL 0.1 M DTT was added to the tube and incubated at 42°C for 2 mins. Finally, 0.5 μL SuperScript II Reverse Transcriptase was added and the solution was incubated 42°C for 50 mins. The Reverse Transcriptase was inactivated by heating at 70°C for 15 min. cDNAs were diluted to a final volume of 60 μL.

Quantitative PCR (qPCR)

qPCR was performed in 20 μL final volume with 10 μL SYBR Green JumpStart Taq ReadyMix (Sigma), 1.5 μL cDNA, 1 μL 10 μM Primer F, 1 μL 10 μM Primer R and 6.5 μL H_2O. qPCR primer sequences can be found in Appendix A. Reactions were run on the iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad) using the program detailed in Table 2.5. Relative expression values were determined using the comparative cycle threshold method ($2^{-\Delta\Delta Ct}$) and Ubox housekeeping gene (AT5G15400) as a reference gene.

Table 2.5: Thermal cycling conditions used for qPCR amplification reactions.

Step	**Temperature (°C)**	**Time**	**Cycles**
Initial denaturation	94	2 min	1
Denaturation	94	15 sec	40
Annealing, elongation and fluorescence measure	62	1 min	
Dissociation curve	40 - 98	10 sec / 0.5 °C	1

2.6 Agilent microarrays

2.6.1 RNA extraction

Total RNAs were extracted as detailed in Section 2.5.7, except RNA was isolated and cleaned using RNeasy Plant Mini Kit (Qiagen) following manufacturer's recommendations. RNA quality was initially checked by 1% agarose gel followed by an integrity check using the Agilent Bioanalyser System (Bioanalyzer RNA 6000 Nano kit). RNA samples were quantified with a NanoDrop ND-1000.

2.6.2 Amplification and labelling

One hundred ng of total RNA was amplified and labelled with Cy3-CTP using the Low Input Quick Amp Labeling Kit (Agilent) following the manufacturer's recommendations. After amplification and labelling reactions, cRNAs were purified

on a RNeasy Mini Spin Column (Agilent) and quantified using a NanoDrop ND-1000.

2.6.3 Hybridisation, washing and scanning

One point six five μg Cy3 labelled cRNAs were fragmented is 25 x Fragmentation Buffer at 60°C for 30 mins. Samples were then put on ice and 2 x Hi-RPM Hybridization Buffer was added to stop the fragmentation reaction. Samples were loaded onto Agilent Arabidopsis (V4) Gene Expression Microarrays (4x44K, G2519F-021169) and hybridised at 65°C for 17 hours.

Arrays were washed in Gene Expression Wash Buffers (with 0.005% Triton X-102 (10%)) following the manufacturer's protocol. Arrays were scanned immediately after washing using the NimbleGen MS200 scanner (NimbleGen, Roche) at 532 nm (Cy3). Agilent Feature Extraction Software was used to align the array template to scanned images and to extract per probe intensity values.

2.6.4 Data normalisation and differential expression

Pre-processing of data (i.e. quality assessment, array normalisation and filtering) was performed in R (Ihaka and Gentleman, 1996). To identify significantly differentially expressed genes (DEGs), the Bioconductor LIMMA (Linear Models for Microarray Data) software package was used to fit linear models to pairs of samples (Ritchie et al., 2015). Transcripts were classified as differentially expressed if they displayed a $\log_2$ fold-change of ≥ 1.5 and a Benjamini-Hochberg adjusted p-value ≤ 0.05. R code to identify differentially expressed genes was written by Alonso Pardal and Anna Gonzalez Gil (fellow PhD students in the group).

2.6.5 GO term and motif enrichment

Venn diagrams were produced using either the online tool Venny (Oliveros, 2007) or the R package VennDiagram (Chen and Boutros, 2011). Over-representation of known transcription-factor binding motifs in promoter sequences was performed using the AME tool from MEME-suite (Bailey et al. (2009); McLeay and Bailey (2010)). To identify over-represented Gene Ontology terms in data, the BiNGO (Biological Network Gene Ontology) Cytoscape plugin (Maere et al., 2005) was used.

2.7 Homology modelling

2.7.1 Template selection and sequence alignment

To identify optimal template structures to use for each homology model, the Protein Data Bank (PDB, (Bernstein et al., 1977)) was searched using protein-protein BLAST (Blastp) (Altschul et al., 1990) with the target protein sequence as query input (substitution matrix BLOSUM62, gap penalty=11 and extend penalty=1). Metrics extracted from the BLAST search included percent identity, percent coverage, E- (expect) value and maximum score. Candidate template PDB entries were found and information about the structure determination method, resolution and co-crystallised ligands were collected.

Following assessment and identification of optimal templates, alignments were made between template and target sequences. EMBOSS Water (Smith-Waterman algorithm) (Rice et al., 2000) was utilised to perform pairwise local alignments with substitution matrix BLOSUM62, gap penalty = 10 and extend penalty = 0.5. Alignments were visualised in Jalview (Waterhouse et al., 2009), ensuring maximum sequence similarity and minimal gaps.

Table 2.6: Protein Data Bank (PDB) structures used in this study.

PDB ID	Protein	Used for
4DNC	HsKAT8	Homology model template
1YGH	ScGCN5	Homology model template
5J8F	HsKAT8	Assess how MSL1 affects HsKAT8 structure
2OU2	HsKAT8	Assess how Ac-CoA affects HsKAT8 structure
		Validation of AutoDock Vina docking algorithm

2.7.2 Model construction and evaluation

MODELLER (Sali and Blundell, 1993) is a freely available program for protein structure modelling and was utilised to create the homology models in this study. Sixty four initial models were generated using the automodel class implemented in MODELLER (v 9.17), followed by 16 loop model optimisations per starting model yielding 1,024 models in total. Optimisations were performed by employing the

maximum level of conjugate gradients and molecular dynamics with simulated annealing as permitted by MODELLER protocols, and these procedures were repeated twice for each model. Modelling computations were performed on a High Performance Computing (HPC) cluster, based on Dell PowerEdge servers and CentOS with 2.5 TB RAM and 224 AMD Opteron cores.

All generated models were ranked according to their Discrete Optimized Protein Energy (DOPE) (Shen and Sali, 2006) and Molecular Potential Density Function (molpdf) (Sali and Blundell, 1993) scores. Restraint violation profiles for the highest ranking models were assessed, and those without heavy violations were identified as candidate models. The model which was the most highly ranked for both DOPE and molpdf scores was identified and underwent further quality assessment.

Evaluation of model quality was extended by comparing the DOPE-per-residue profile of the model against the template structure. Several structural validation steps were also performed including Ramachandran plot analysis, which identifies steric clashes within the polypeptide backbone (Lovell et al., 2003), and ProSA analysis (Wiederstein and Sippl, 2007). ProSA is a program that calculates several energy-based protein quality scores (Sippl, 1993) and plots them in the context of all known protein structures.

2.8 Compound library preparation

ChemBridge screening libraries were retrieved as SD (Structure Data) files containing compound atom types and coordinates (amongst other descriptors). Two non-overlapping libraries were used in the virtual screening experiments: CORE Library stock (approx. 620,000 total structures) and EXPRESS-Pick Collection stock (approx. 460,000 total structures). Initially, the virtual screen began with more than 1 million unique structures that could be individually selected and purchased.

All SD files were imported into the compound management software DataWarrior (Sander et al., 2015) for visual inspection and filtering purposes. To convert compound SD files into the appropriate file format for *in silico* screening, OpenBabel (O'Boyle et al., 2011) was utilised. During this step, SD files were converted to PDB format with 3D coordinates and hydrogen atoms assigned appropriate for protonation states at pH 7.

The docking algorithm employed for this study, AutoDock Vina (Trott and Olson, 2010), requires ligand files to be written in PDBQT format. This format is essentially the same as PDB, except partial charges ('Q') and AutoDock atom types ('T') are included. The graphical user interface (GUI) Raccoon v 1.0b (Forli et al., 2016) was used to perform this conversion for all compounds tested in the screen.

2.9 *In silico* docking

Homology models were generated in PDB file format and, like the ligand files, were converted to PDBQT format using the GUI for AutoDock, AutoDock Tools (Morris et al., 2009). The parameters of the docking grid-box were also determined using AutoDock Tools. AutoDock Vina was run with exhaustiveness = 16 and the virtual screen was run on a High Performance Computing (HPC) cluster, based on Dell PowerEdge servers and CentOS with 2.5 TB RAM and 224 AMD Opteron cores.

Chapter 3

HAM2 is a Moderator of Plant Immunity and Growth

3.1 Introduction

As described in Section 1.3.6, transcriptional reprogramming is a fundamental process in the co-ordination of defence responses (Jenner and Young, 2005). Multiple publications have described the differential expression of $\approx$ 10% of the *Arabidopsis* genome after exposure to the bacterial pathogen-associated molecular pattern (PAMP) flg22 (Zipfel et al., 2004; Denoux et al., 2008; Frei dit Frey et al., 2014). Flg22 is a 22-amino acid fragment derived from bacterial flagellin which stimulates the FLS2 receptor to initiate downstream PAMP-triggered immunity (PTI) responses (Gómez-Gómez and Boller, 2000). After performing microarray experiments on 13 day old seedlings after exposure to flg22 for 30 minutes, Frei dit Frey et al. (2014) found that expression of 1,529 genes was induced and expression of 862 genes was repressed (indicating an overall increase in gene expression) following flg22 elicitation. These transcriptomic changes were shown to be dependent on flg22 perception by Zipfel et al. (2004) who also analysed the flagellin-insensitive mutant *fls2-17* and found minor changes in only 6 genes after treatment with flg22.

In Section 1.4, mechanisms underlying gene expression regulation were described. The transcription of a gene is highly dependent on the chromatin structure at each loci, which in turn is regulated by developmental cues and external stimuli. Chromatin structure in specific regions is tightly regulated by histone-modifying enzymes and ATP-dependent chromatin remodelling complexes (Wegel and Shaw, 2005; Jerz-

manowski and Archacki, 2017). Histone-modifying enzymes can catalyse the transfer of methyl, acetyl or phospho groups to the N-terminus of histone tails. These modifications can directly affect chromatin structure by changing DNA/histone interactions within the nucleosome or indirectly by recruiting histone-modifying complexes (Kouzarides, 2007; Hsieh and Fischer, 2005).

In the context of plant immunity, where it is known that 10% of the *Arabidopsis* genome is differentially expressed upon initiation of PTI, it is plausible to hypothesise that histone-modifying enzymes (such as histone acetyltransferases) function during elicitation of defence responses. Since histone acetyltransferases (HATs) are key regulators of gene transcription (Kouzarides, 2007), it was therefore hypothesised that HATs play a role in establishing and regulating responses against pathogens.

In previous work within the group, a reverse genetics approach was taken to assess the response of *Arabidopsis hat* mutants to the hemibiotrophic bacterial pathogen *Pseudomonas syringae* pv. *tomato* (*Pto*) DC3000. T-DNA insertion SALK lines were obtained from the Nottingham Arabidopsis Stock Centre (NASC) (Alonso et al., 2003), and one homozygous line for 10 of the 12 *Arabidopsis* HATs were tested for altered defence responses (it was not possible to obtain homozygous lines for the remaining 2 HATs). Plant lines were inoculated with *Pto* DC3000 and bacterial growth was quantified 3 days post-inoculation. The flagellin-insensitive mutant *fls2* is known to be more susceptible than Col-0 to *Pto* DC3000 (Zipfel et al., 2004) so was included as an experimental control. Of the 10 *hat* mutants tested, one line (*hag1-6*) displayed increased susceptibility whilst another line (*ham2-75*) showed enhanced resistance to *Pto* DC3000 infection (Fig. 3.1). These results indicate that *Arabidopsis* HATs have divergent roles in the regulation of immunity and invited further investigations, particularly into the MYST-family member HAM2 and the GNAT-family member HAG1. Whilst the current chapter focuses on HAM2, the role of HAG1 in plant immunity in investigated in Chapter 5 of this thesis.

As described in Section 1.5.1, HAM2 belongs to the MYST family of *Arabidopsis* HATs. Using *in vitro* assays, both HAM1 and HAM2 have been shown to specifically target histone 4 lysine 5 (H4K5) for acetylation (Earley et al., 2007). In contrast to acetylation of H3 tail residues, which is associated with actively transcribed regions, the function of acetylation of H4 lysine residues is not so clear. To

date, there have been no publications describing whether H4K5ac is associated with transcriptional activation or repression in *Arabidopsis*; it is therefore possible that HAM1 and HAM2 act as transcriptional activators or repressors. Despite this, we may glean information from MYST-family homologues in other species. The MYST HATs in human (KAT5 (Lysine Acetyltransferase 5)), *S. cerevisiae* (SAS2 (Something About Silencing)) and *Drosophila* (Chm (Chameau)) have been shown to be required for the transcriptional repression at specific genes, and also for maintaining the structure of transcriptionally silent chromatin (Zou and Bi, 2008; Grienenberger et al., 2002; Xiao et al., 2003). In *Arabidopsis*, several publications have found that HAM1 and HAM2 function redundantly in developmental processes (Latrasse et al., 2008) and response to UV-B induced DNA damage (Campi et al., 2012), but a role for MYST HATs has not yet been published in plant defence responses.

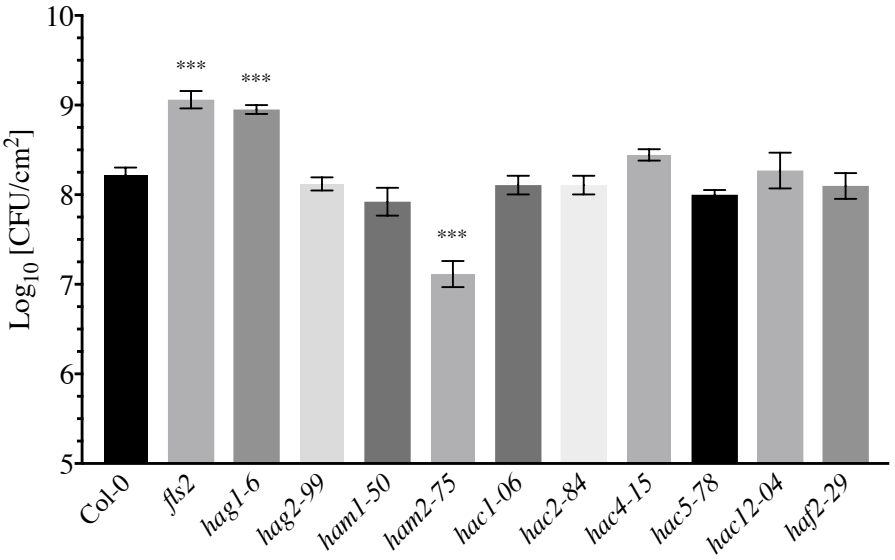


Figure 3.1: HAT mutants *ham2-75* and *hag1-6* are more resistant and more susceptible to *Pto* DC3000 respectively. *fls2* was included as a susceptible control. Five week old plant lines were sprayed with *Pto* DC3000 (OD_{600}= 0.1) and bacterial growth in leaves was quantified 3 days post-inoculation. Data shown are representative of 3 independent experiments. Statistical significance versus Col-0 determined by two-tailed t-test, n = 6, *** $P \leq 0.001$. Error bars indicate standard error. Genotyping for homozygosity was performed by Ntiana Mamafidou (an Erasmus student) and the bacterial growth assay was performed by Sophie Piquerez (a post-doctoral research assistant).

In preliminary investigations into the role of HAM2 in plant immunity, the expression of *HAM2* was analysed in published time-course transcriptomic datasets of *Arabidopsis* ecotype Col-0 inoculated with the necrotrophic fungus *Botrytis cinerea* ((Windram et al., 2012), Gene Expression Omnibus (GEO) accession number: GSE29642), *Pto* DC3000 and the strain *Pto* DC3000*hrpA*- ((Lewis et al., 2015), GEO accession number: GSE56094). *Pto* DC3000*hrpA*- is unable to deliver effectors into the host cell due to disruption of type III secretion system assembly (Wei et al., 2000). As seen in Figure 3.2A, *HAM2* expression is down-regulated in Col-0 by approximately two-fold 26 hours post-infection with *Botrytis cinerea*. Similarly, but to a lesser extent, *HAM2* is also down-regulated following *Pto* DC3000 infection, particularly for DC3000*hrpA*-, from 3 hours post-infection onwards (Fig. 3.2B). This transcription profile of *HAM2* following infections with *B. cinerea*, *Pto* DC3000 and *Pto* DC3000*hrpA*-, together with findings from the reverse genetic screen (Fig. 3.1), supports the hypothesis that *HAM2* functions as a negative regulator of basal plant defence responses against both hemibiotrophs and necrotrophs.

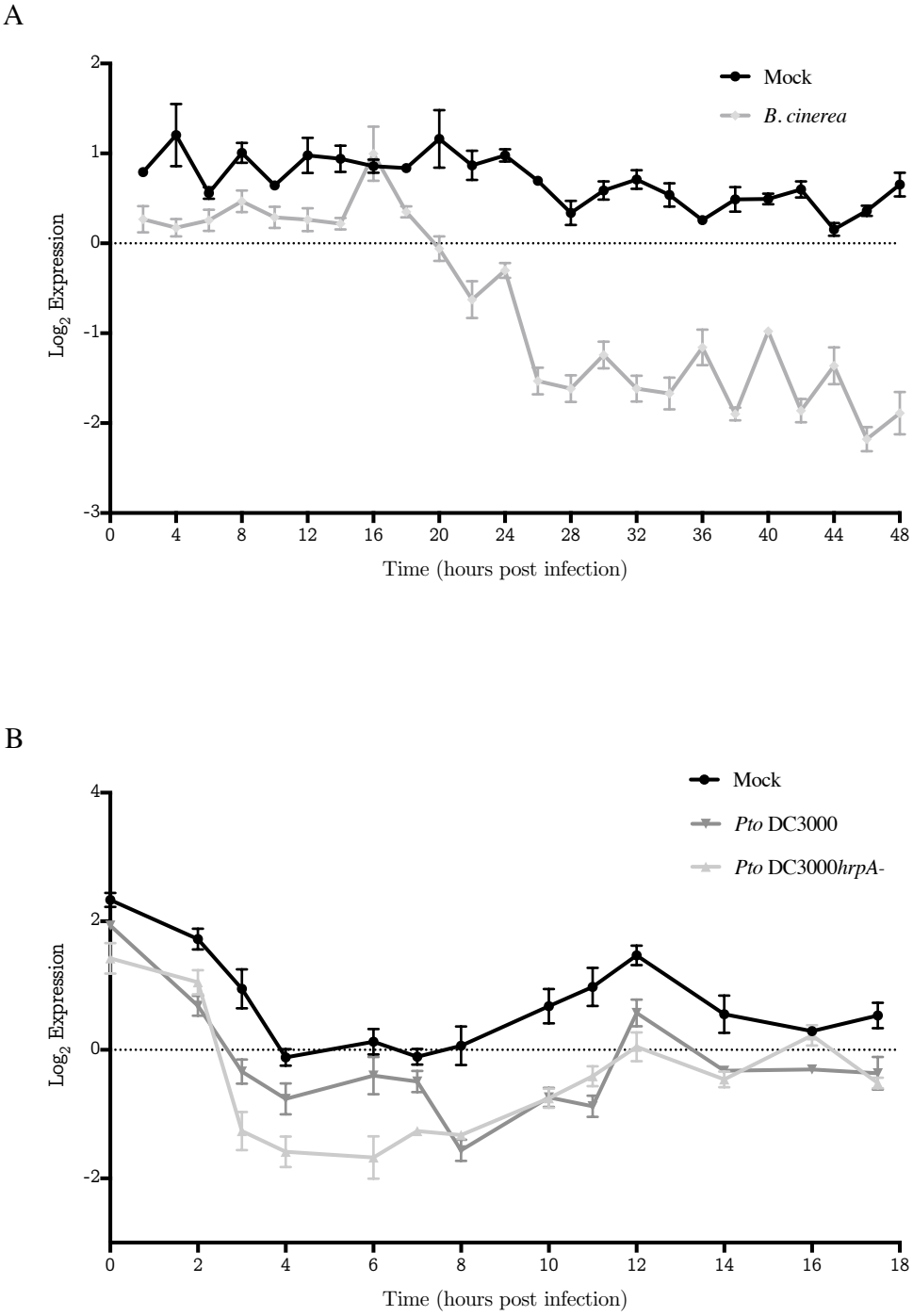


Figure 3.2: Expression of *HAM2* (AT5G09740) in response to infection with (A) *Botrytis cinerea* (as published by Windram et al. (2012), GEO accession number: GSE29642) and (B) *Pto* DC3000 and DC3000*hrpA*- (as published by Lewis et al. (2015), GEO accession number: GSE56094). Data points represent the mean of four biological replicates and error bars indicate standard error.

3.1.1 Aims of this chapter

The aims of this chapter are to establish whether mutation of *HAM2* in *Arabidopsis* alters:

1. Resistance/susceptibility to *Pto* DC3000 infection, a representative hemibiotrophic bacteria
2. Responses elicited by the bacterial PAMP flg22
3. Effector-triggered immunity responses
4. Resistance/susceptibility to *Botrytis cinerea* infection, a representative necrotrophic fungi
5. *Arabidopsis* growth and development phenotypes

3.2 HAM2 is a negative regulator of immunity

3.2.1 *ham2* mutants show enhanced resistance to *P. syringae* infection

To ensure that the previously observed resistance phenotype of *ham2-75* was due to specific mutation of *HAM2*, a second homozygous T-DNA insertion line (*ham2-71*) was obtained. As presented in Figure 3.3B, the insertion within the *HAM2* gene body in *ham2-75* causes a knock-out effect on *HAM2* expression, whereas the insertion in the promoter region, *ham2-71*, causes a knock-down effect. Despite the difference in *HAM2* expression, both mutant lines show a similar resistance to *Pto* DC3000 phenotype (Fig. 3.3C).

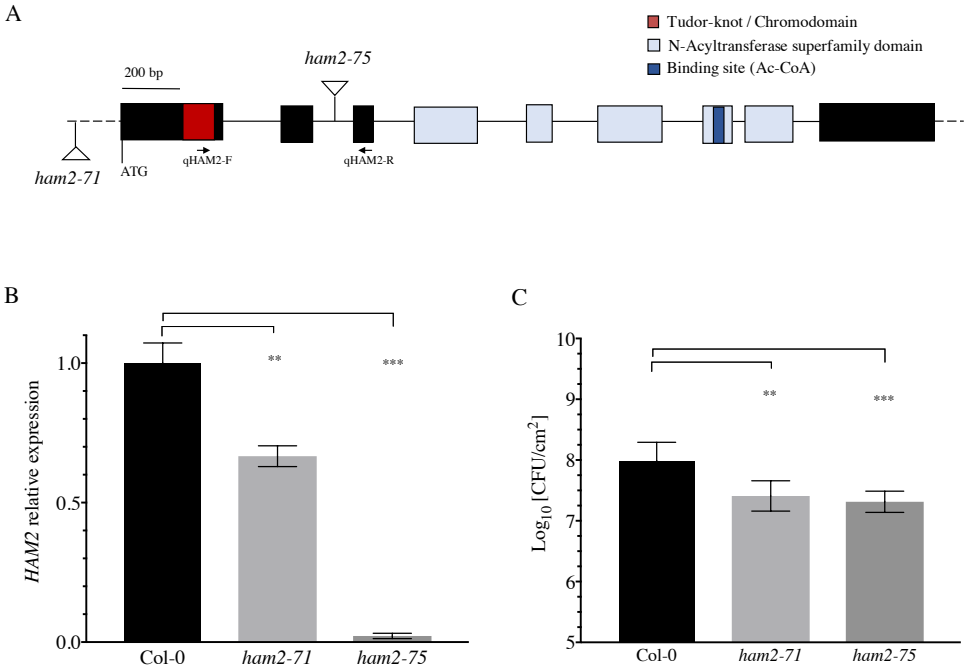


Figure 3.3: Two *ham2* mutant lines are resistant to *Pto* DC3000. (A) T-DNA insertion positions along the *HAM2* gene. Bars indicate position of exons, lines indicate introns and untranslated regions. (B) Relative expression of *HAM2* in Col-0, *ham2-71* and *ham2-75* in 4 week-old plants in resting conditions was determined by qPCR using gene-specific primers (Section 2.5.7). (C) Growth of *Pto* DC3000 in *ham2* lines (Section 2.3.2). 4-5 week old plant lines were sprayed with *Pto* DC3000 ($OD_{600} = 0.1$) and bacterial growth in leaves was quantified 3 days post-inoculation. Data shown are representative of 3 independent experiments. Statistical significance versus Col-0 determined by two-tailed t-test, n = 6, ** $P \leq 0.01$, *** $P \leq 0.001$. Error bars indicate standard error.

3.2.2 Microarrays following flg22 treatment

To investigate whether the enhanced resistance to *Pto* DC3000 phenotype of *ham2* mutants was due to transcriptional differences during early PTI responses, microarrays were performed on Col-0 and *ham2-75* (knock-out mutant) prior to and following elicitation with flg22. By eliciting with flg22 only, we can be confident that PTI responses only will be analysed in this experiment, since infection with *Pto* DC3000 would also involve the delivery of effectors into the host. RNA samples were extracted from pooled 16 day old Col-0 and *ham2-75* seedlings with or without 1 hour elicitation with 1 μM flg22 and samples were hybridised onto Agilent *Arabidopsis* (V4) microarrays containing 43,803 probes. Three biological replicates per condition were performed on separate occasions. The average probe intensity was taken for each probes across replicates post array normalisation. A full description of experimental procedures and data processing can be found in Section 2.6. Experimental procedures were carried out by Sophie Piquerez (post-doctoral research assistant).

Quantile array normalisation was performed with the Bioconductor LIMMA (Linear Models for Microarray Data) software package in R (Ritchie et al., 2015). To ensure the quality of the data, box-plots were produced to indicate the overall intensity of arrays. These are presented before normalisation and after normalisation in Figure 3.4. It can be seen that the data normalisation procedure was successful as indicated by the alignment of all box-plots.

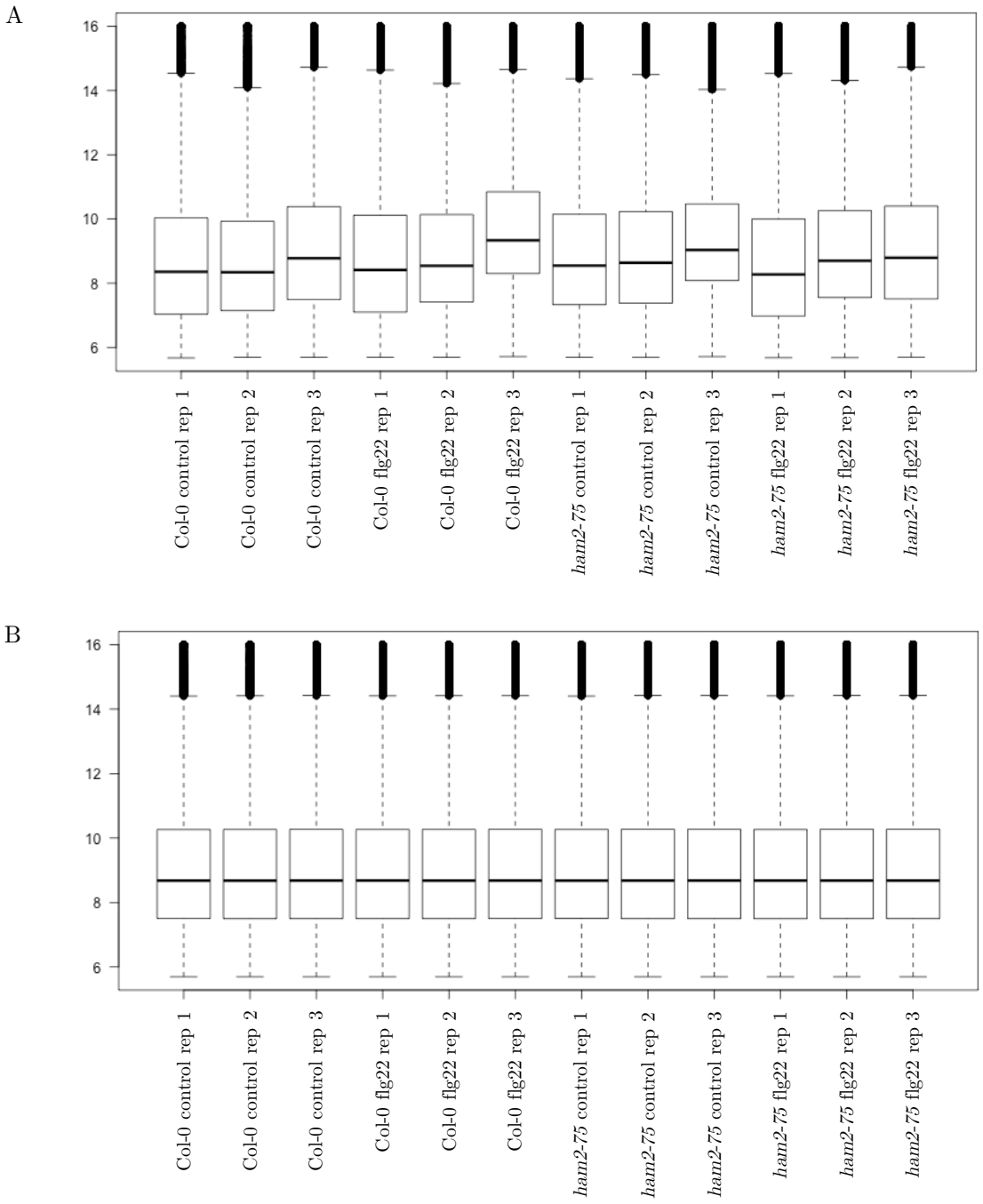


Figure 3.4: Overall microarray intensities before and after quantile normalisation. (A) Array intensities before normalisation. (B) Array intensities after normalisation. The centre of the box indicates the median, the box length indicates the interquartile range (IQR, 3rd quartile − 1st quartile), the upper whisker indicates the 3rd quartile + 1.5 * IQR, whereas the lower whisker indicates the 1st quartile − 1.5 * IQR of the data.

Initial analysis of differentially expressed genes (DEGs) indicated that results from this microarray experiment are in agreement with those published from similar experiments (Denoux et al., 2008; Frei dit Frey et al., 2014). Using the thresholds of adjusted p-value ≤ 0.05 and $\log_2$ fold-change of ≥ 1.5, 2881 genes were up-regulated and 1688 genes were down-regulated in Col-0 after 1 hour of elicitation with flg22. In *ham2-75*, 3092 genes were up-regulated and 1782 genes were down-regulated after flg22 elicitation. The differential expression between the four conditions is represented as a heat map in Figure 3.5. The majority of changes can be observed between treated and untreated samples, although some differences between Col-0 and *ham2-75* in untreated samples can also be seen.

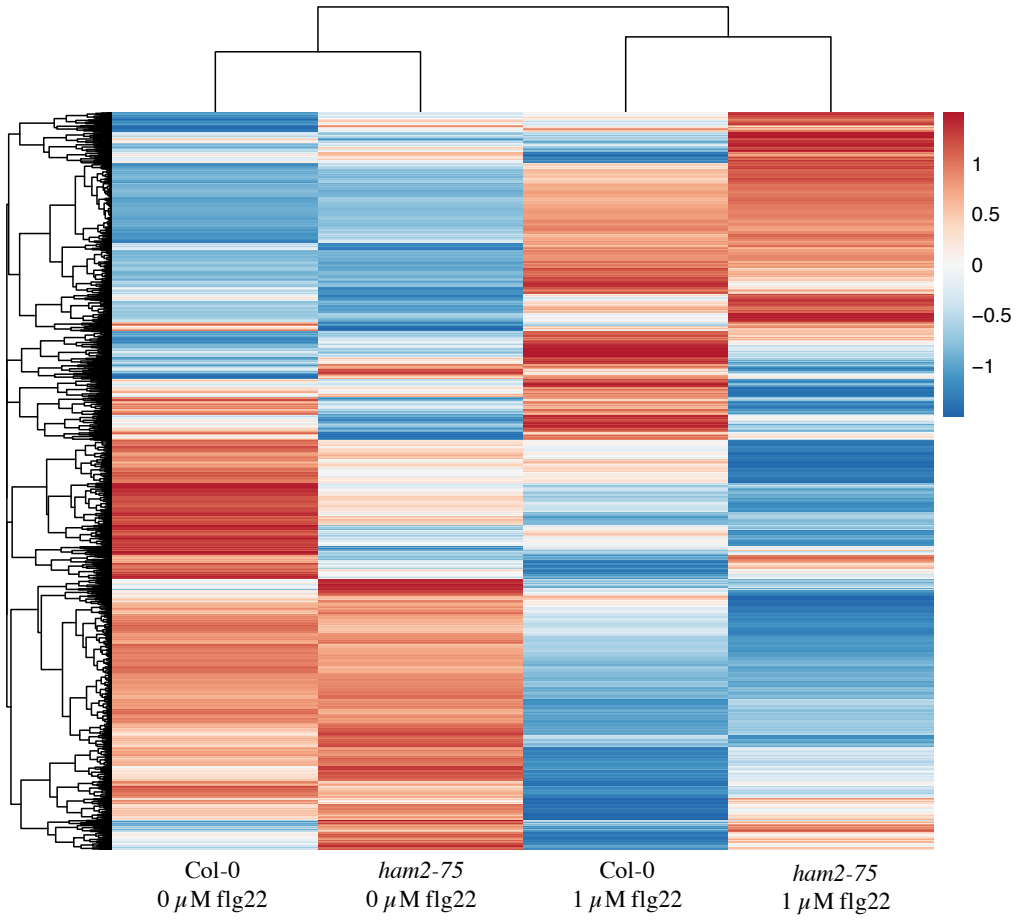


Figure 3.5: Heat map representation of differential gene expression in Col-0 and *ham2-75* prior to and following flg22 elicitation. Relative expression values after array normalisation were subjected to hierarchical clustering. Red indicates high expression and blue indicates low expression.

Differentially expressed genes without flg22 elicitation

To analyse differences between genotypes, differential expression analysis was performed between Col-0 and *ham2-75* without flg22 elicitation. Based on thresholds of p-value ≤ 0.05 and $\log_2$ fold-change of ≥ 1.5, 6 genes were found to be differentially expressed between genotypes. This small number of differentially expressed genes indicates that the transcriptomes of Col-0 and *ham2-75* are very similar under basal conditions at the seedling stage when this experiment was performed.
The genes with greater expression in *ham2-75* vs Col-0 were:

- *WHY2* (AT1G71260), involved in DNA repair, defence response, regulation of transcription
- AT1G76405, an outer envelope pore 21B-like protein
- *PR5K* (AT5G38280), involved in response to fungus
- AT5G26270, a transmembrane protein

The genes with reduced expression in *ham2-75* vs Col-0 were:

- *NEAP2* (AT5G26770), a nuclear envelope protein
- *QQS* (AT3G30720), involved in negative regulation of starch metabolic process, positive regulation of protein metabolic process

Flg22 up-regulated genes in *ham2-75* exclusively

The results presented previously in this chapter (Figs. 3.2 and 3.3) support a model in which HAM2 functions as a negative regulator of plant immunity, since *ham2* mutants are more resistant to *Pto* DC3000 and *HAM2* is down-regulated upon pathogen infection. With the aim of identifying genes that are under the repressive action of HAM2 in Col-0, but not in the *ham2-75* mutant after flg22 elicitation, an analysis of genes up-regulated in response to flg22 in *ham2-75*, was performed. Also, HATs regulate gene transcription by modulating the accessibility of chromatin of specific region and altering local chromatin structure can affect both the timing and amplitude of transcriptional responses to stimuli. It was hypothesised that, since *ham2-75* is a histone acetyltransferase null mutant, phenotypic differences seen between Col-0 and *ham2* mutants may be due to differences in the degree of gene up-regulation, i.e. a defence gene may be up-regulated in both genotypes but more so in *ham2-75*. Therefore, the expression levels of genes that were up-regulated in both Col-0 and *ham2-75* under 1 μM flg22 conditions were also examined.

Firstly, to analyse the genes up-regulated following flg22 elicitation in *ham2-75* exclusively, the following steps were carried out:

1. Identify genes that are up-regulated after 1 μM flg22 elicitation in *ham2-75* but not Col-0
2. Perform analyses to identify enriched GO terms. This method makes use of the Gene Ontology system of classification where genes are categorised according to their functional characteristics. A statistical test is carried out to assess if a specific GO term category is enriched within a gene-set when compared to a reference group
3. Perform analysis to identify over-represented known transcription factor (TF) binding motifs in the promoters of these genes. This method was used to determine if more highly expressed genes in *ham2-75* (i.e. under the repressive action of HAM2 in Col-0) contain any motifs to which HAM2 activity is directed to.

Analysis of DEGs indicated that 322 genes were up-regulated following treatment with flg22 exclusively in *ham2-75* (Fig. 3.6A). To determine whether the enhanced expression of these genes contributes to the observed *Pto* DC3000 resistance phenotype of *ham2-75*, the enrichment for specific GO terms and TF-binding motifs within this set was assessed. This analysis found that no GO terms were significantly enriched when the reference population was the *Arabidopsis* Col-0 genome (TAIR10) and also when the reference population was the set of genes up-regulated after flg22 with equal expression levels between Col-0 and *ham2-75*. Similar results were obtained from motif analysis also; no known TF-binding motifs were over-represented in the promoters of genes exclusively up-regulated in *ham2-75*.

Flg22 up-regulated genes in Col-0 and *ham2-75*

As mentioned previously, since *ham2-75* is a histone acetyltransferase null mutant, we hypothesised that phenotypic differences seen between Col-0 and *ham2-75* may be due to differences in the degree of gene up-regulation. Based on this, the expression levels of genes that were up-regulated in both Col-0 and *ham2-75* under 1 μM flg22 conditions were analysed. For this analysis, the following steps were carried out:

1. Identify genes that are up-regulated after treatment with 1 μM flg22 in both Col-0 and *ham2-75*

2. From the up-regulated genes, identify those with higher expression values in *ham2-75*

3. Perform enrichment analysis for specific GO terms or TF-binding motifs on the up-regulated genes with greater expression values in *ham2-75*

4. Analyse the expression profile of these genes in time-course transcriptomic datasets following inoculation with *Pto* DC3000 and DC3000*hrpA*- ((Lewis et al., 2015), GEO accession number: GSE56094)

A set of 2025 genes were identified as being up-regulated following treatment with flg22 in both Col-0 and *ham2-75* (Fig. 3.6A). The relative expression values of this set of 2025 genes across the different conditions are presented in heat map form in Figure 3.6B. The up-regulation of these genes in flg22-treated samples is confirmed for both genotypes, and differences in relative expression levels between Col-0 and *ham2-75* in non-treated samples can also be observed.

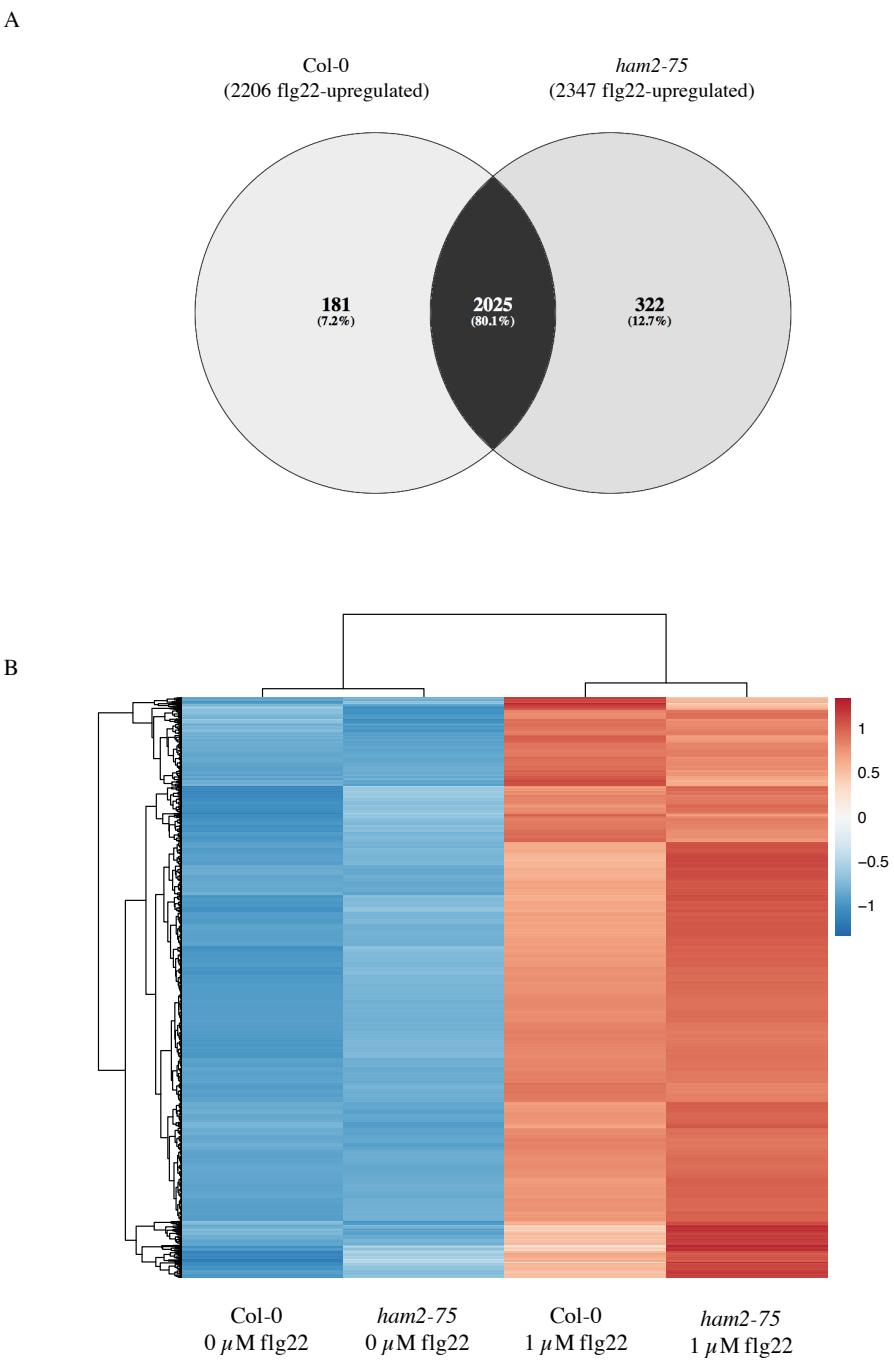


Figure 3.6: Identified flg22 up-regulated genes in Col-0 and *ham2-75*. (A) The overlap of genes up-regulated in both Col-0 and *ham2-75* following flg22 treatment demonstrated by Venn diagram. In Col-0, 2206 genes were up-regulated and in *ham2-75*, 2347 genes were up-regulated between control (0 μM flg22) and 1 μM flg22 conditions. (B) Heat map representation of relative expression values of the 2025 set of genes up-regulated in both Col-0 and *ham2-75* under control (0 μM flg22) and 1 μM flg22 conditions. Data shown are relative expression values after array normalisation. Red indicates high expression and blue indicates low expression.

To identify genes within this set with higher relative expression in *ham2-75* compared to Col-0, the difference between the expression of each gene in Col-0 and *ham2-75* after treatment with flg22 was calculated. These data are presented in histogram form in Figure 3.7. The data form a normal distribution with a mean value of 0.6, indicating that there is a trend for expression values of flg22-responsive genes to be higher in *ham2-75* than Col-0. The green line in this figure indicates zero difference in relative expression values: 436 genes to the left of this line have lower expression in *ham2-75* than Col-0, 1589 genes to the right of this line have higher relative expression values in *ham2-75*. The fact that there are more flg22 up-regulated genes with higher expression values in *ham2-75* suggests that HAM2 does indeed function as a transcriptional repressor. In order to focus on the genes that are distinctively more up-regulated in *ham2-75* than Col-0, a threshold of $\log_2$ difference of ≥ 1.5 was established (red dashed line in Fig. 3.7). Based on this threshold, a set of 114 genes were identified as having heightened expression in *ham2-75* than Col-0 after exposure to flg22.

GO term enrichment in genes more highly expressed in *ham2-75* than Col-0 post-flg22

To analyse the properties of this gene-set more deeply, 'Gene Ontology' (GO) term analysis was performed on the 114 flg22-responsive genes with enhanced expression in *ham2-75*. GO terms are used to characterise subsets of genes based on associated biological processes, molecular functions or cellular components. BINGO (Biological Network Gene Ontology) (Maere et al., 2005) is Cytoscape plug-in that identifies statistically over-represented GO terms within a set of genes. With the *Arabidopsis* Col-0 genome (TAIR10; 28,775 genes not including transposable elements or pseudogenes) set as the reference population, the gene-set was enriched in the GO terms presented in Table 3.1. The resulting GO terms include expected findings such as protein phosphorylation, defence response and response to stimulus.

GO term analysis was then performed on the same set of 114 genes, but with the reference population set as genes up-regulated after flg22 with equal expression levels between Col-0 and *ham2-75*. In this way we could uncover a specific aspect of responses to flg22 that are controlled by HAM2. Results from this search indicated that, amongst flg22-responsive genes, there are no GO terms that are enriched specifically in genes more highly expressed in *ham2-75* compared to Col-0.

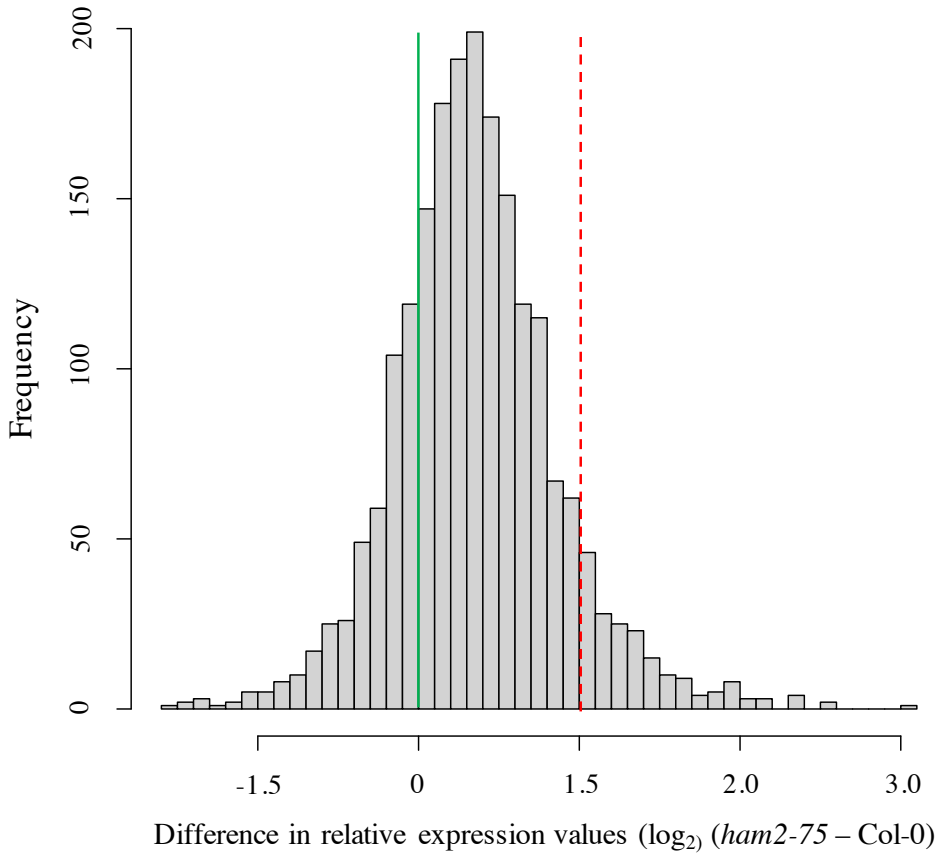


Figure 3.7: The difference in expression values of genes up-regulated after exposure to flg22 in Col-0 and *ham2-75*. The green line indicates zero difference in relative expression values. Genes to the left of this line have lower expression in *ham-75* than Col-0, genes to the right of this line have higher expression in *ham2-75* than Col-0. The red dashed line indicates where the $\log_2 \geq 1.5$ difference threshold was applied.

TF-binding motif enrichment analysis in genes more highly expressed in *ham2-75* than Col-0 post-flg22

The promoters of the 114 flg22-responsive gene-set with enhanced expression in *ham2-75* were then analysed to identify over-represented known TF-binding motifs. Two *Arabidopsis* motif databases were searched: 1. *Arabidopsis* PBM database containing 113 TF-binding motifs identified by Franco-Zorrilla et al. (2014) using protein-binding microrarrays (PBM) and 2. *Arabidopsis* DAP-seq database containing 872 TF-binding motifs identified by O'Malley et al. (2016) using DNA affinity purification sequencing (DAP-seq). The regions 500 bp upstream from the transcription start site of each of the 114 flg22-responsive genes were analysed with the

Table 3.1: GO terms significantly enriched in the 114 flg22-responsive genes with enhanced expression in *ham2-75*. Here, the reference population gene set was the *Arabidopsis* Col-0 genome (TAIR10; 28,775 genes not including transposable elements or pseudogenes). Hypergeometric test with Benjamini and Hochberg FDR correction, threshold p-value ≤ 0.05.

	Term	**Observed frequency**	**p-value**
GO:0010200	response to chitin	7 / 103 genes, 6.8%	0.000316
GO:0006468	protein phosphorylation	15 / 103 genes, 14.6%	0.000482
GO:0016310	phosphorylation	15 / 103 genes, 14.6%	0.000482
GO:0050896	response to stimulus	33 / 103 genes, 32%	0.000482
GO:0006793	phosphorus metabolic process	15 / 103 genes, 14.6%	0.000568
GO:0006796	phosphate metabolic process	15 / 103 genes, 14.6%	0.000568
GO:0006952	defense response	13 / 103 genes, 12.6%	0.000568
GO:0009743	response to carbohydrate stimulus	7 / 103 genes, 6.8%	0.000753
GO:0006464	protein modification process	16 / 103 genes, 15.5%	0.00281
GO:0043412	macromolecule modification	17 / 103 genes, 16.5%	0.00296
GO:0006874	cellular calcium ion homeostasis	3 / 103 genes, 2.9%	0.00474
GO:0055074	calcium ion homeostasis	3 / 103 genes, 2.9%	0.00532
GO:0072503	cellular divalent inorganic cation homeostasis	3 / 103 genes, 2.9%	0.0054
GO:0006950	response to stress	20 / 103 genes, 19.4%	0.00595
GO:0072507	divalent inorganic cation homeostasis	3 / 103 genes, 2.9%	0.00782

background set as genes up-regulated after flg22 with equal expression levels between Col-0 and *ham2-75*. The region 500 bp upstream of the transcription start site was analysed because very low nucleosome occupancy is found in this region of promoters of protein-coding genes, and nucleosome occupancy is closely correlated with transcription levels (Li et al., 2014b).

The results from searches against both databases showed an enrichment for WRKY TFs. From the Franco-Zorrilla et al. (2014) PBM database, WRKY18, WRKY38, WRKY12 and WRKY45 were enriched. From the O'Malley et al. (2016) DAP-seq database, WRKY18, WRKY40, WRKY70, AT3G42860 (zinc knuckle (CCHC-type)

family protein), WRKY30, WRKY47 and WRKY55 were enriched (Table 3.2). The most significant hit from both searches was WRKY18, although other WRKYs were also identified that were different between the two searches (most likely due to the differences between the contents of the databases searched).

Table 3.2: Over-represented known transcription factor binding motifs in genes which are more highly expressed in *ham2-75* than Col-0 after flg22 treatment. Here, the reference population was set to genes up-regulated after flg22 with equal expression levels between Col-0 and *ham2-75*. One-tailed Wilcoxon rank-sum test was used, threshold p-value ≤ 0.05.

Database	ID	Motif	Adjusted p-value
PBM	WRKY18		4.20E-05
	WRKY38		3.38E-04
	WRKY12		9.07E-04
	WRKY45		9.11E-04
DAP-seq	WRKY18		5.65E-04
	WRKY40		3.07E-03
	WRKY70		4.13E-03
	AT3G42860		4.29E-03
	WRKY30		7.97E-03
	WRKY47		7.97E-03
	WRKY55		8.71E-03

WRKY18 expression in *ham2-75*

Since WRKY18 was the most significantly enriched motif from both database searches, the potential relationship between HAM2 and WRKY18 was explored further. Previous analysis discovered that genes containing the WRKY18-binding motif in the promoter regions were found to be more highly expressed in *ham2-75* lines than Col-0 after flg22 treatment. An initial hypothesis explaining this observation was that, in Col-0 plants, HAM2 negatively regulates expression of the *WRKY18* TF itself (either directly or indirectly). In *ham2-75* lines, the negative regulatory effect of HAM2 would be absent so the targets of WRKY18 would be more highly expressed. After assessing the expression of *WRKY18* in Col-0 and *ham2-75*, this hypothesis was disproved as its expression was unchanged between flg22-treated and control conditions in both lines. Therefore, HAM2 does not target the expression of WRKY18 itself.

Expression of genes containing WRKY18-binding motifs following infection with *Pto* DC3000

To investigate potential links between the reduced expression of *HAM2* in response to *Pto* DC3000 (Fig. 3.2B) and the increased expression of WRKY18-motif containing genes following flg22 treatment in *ham2-75*, the expression of WRKY18-motif containing genes was analysed in time-course transcriptomic data following infection with *Pto* DC3000 (Lewis et al. (2015), GSE56094). Firstly, the frequency of the WRKY18-binding motif in the set of 114 flg22-responsive genes with enhanced expression in *ham2-75* was assessed. The motif was found to occur one or more times in the promoter regions of 33 of the 114 flg22-responsive genes with enhanced expression in *ham2-75*. Also, Birkenbihl et al. (2016) recently established the genome-wide binding sites of WRKY18 after treatment with 1 μM flg22 by ChIP-seq analysis. The list of 33 WRKY18-binding motif containing genes was compared to the list of WRKY18-bound genes after flg22 treatment, as published by Birkenbihl et al. (2016). Promoters of 16 out of the 33 genes were experimentally validated as bound by WRKY18 after flg22 treatment in 12 day old seedlings (Table 3.3).

Table 3.3: Sixteen of the 114 flg22-responsive genes with enhanced expression in *ham2-75* that contain the WRKY18 motif and are targeted by WRKY18 after treatment with 1 M flg22 as determined by Birkenbihl et al. (2016). Hpi = hours post infection.

WRKY18 motif Gene ID	**Gene Description**	**Expression in *Pto* DC3000 time-course**
AT1G79680	Wall Associated Kinase-Like 10 (WAKL-10)	Upregulated 1.5-fold 2 hpi
AT4G28460	Not known	Upregulated 1.5-fold 4 hpi
AT4G04480	F-box family protein	Not clear
AT4G23170	Receptor-like kinase-related (EP1)	No change
AT3G12910	NAC domain transcriptional regulator protein	No change
AT4G14450	Not known	Upregulated 1.5-fold 2 hpi
AT3G58490	Phosphatidic acid phosphatase (PAP2) family protein	No change
AT1G59590	Not known	Upregulated 1.5-fold 4 hpi
AT4G11300	Not known	Not clear
AT3G19580	Zinc-finger protein 2 (ZF2)	Upregulated 1.5-fold 8 hpi
AT1G74360	Leucine-rich repeat kinase family protein	Upregulated 1.5-fold 2 hpi
AT2G47550	Plant invertase/pectin methylesterase inhibitor superfamily	Upregulated 1.5-fold 2 hpi
AT1G51920	Not known	Upregulated 1.5-fold 2 hpi
AT3G26500	Plant intracellular Ras group-related LRR 2 (PIRL2)	Upregulated 1.5-fold 2 hpi
AT3G19010	2-oxoglutarate and Fe-dependent oxygenase superfamily protein	Upregulated 1.5-fold 2 hpi
AT5G64905	Elicitor peptide 3 precursor (PROPEP3)	Upregulated 1.5-fold 4 hpi

The expression profiles of these 16 genes were then analysed in time-course transcriptomic datasets following infection with *Pto* DC3000 (Lewis et al. (2015), GEO accession number: GSE56094). An up-regulation correlating with the down-regulation of *HAM2* could be seen for 11 genes over the infection time-course. For the remaining 5 genes, an unchanged or unclear expression pattern was observed. Expression profiles for 3 WRKY18-targeted genes which are up-regulated after *Pto* DC3000 infection are presented in Figure 3.8 as examples. The expression profiles of the remaining 13 genes can be found in Appendix B.

In this analysis, the transcriptomes of Col-0 and *ham2-75* were found to be very similar under basal conditions. By analysing genes that were differentially expressed after flg22 elicitation in Col-0 and *ham2-75*, a relationship between WRKY18-motif containing genes and HAM2 was discovered. Although direct causal links cannot be made between HAM2 and WRKY18-motif containing genes in this expression analysis, it provides information on how *ham2* lines achieve heightened resistance to *Pto* DC3000 by directly or indirectly affecting the expression of defence-related genes. As well as WRKY18-motif containing genes, other WRKY-targeted genes identified in the motif analysis (Table 3.2) may also contribute to the resistance phenotype seen for *ham2*.

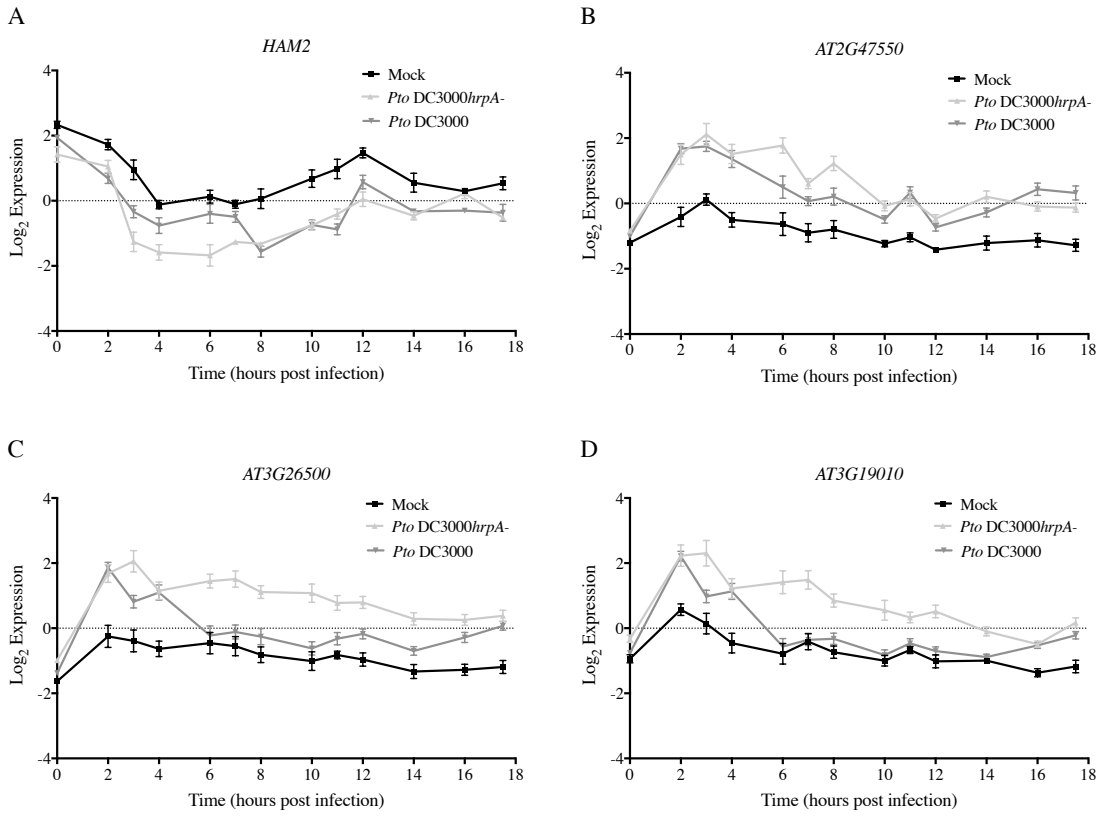


Figure 3.8: Expression of *HAM2* (AT5G09740) and WRKY18 binding motif containing genes in response to infection with *Pto* DC3000 and *Pto* DC3000*hrpA*- (as presented by Lewis et al. (2015), GEO accession number: GSE56094). (A) *HAM2* (AT5G09740), (B) AT2G47550, (C) AT3G26500, (D) AT3G19010.

3.2.3 Effector-triggered immunity is not compromised in *ham2* mutants

The data presented so far in this chapter have indicated that HAM2 has a role in basal defence following infection with *Pto* DC3000. As presented in Section 1.3, a second phase of heightened immune responses are initiated upon host perception of a pathogen effector (effector-triggered immunity). To investigate the possibility of *HAM2* also functioning in ETI, ion leakage assays were performed (described in Section 2.3.4). In this method, hypersensitive response (HR)-dependent cell death following effector recognition is quantitated by measuring electrolyte leakage from infected leaves, as performed by Mackey et al. (2002). Col-0, *ham2-71* and *ham2-75* mutant lines were infected with *Pto* DC3000 expressing the recognised type III effector AvrRpt2. This is a well-studied effector that is recognised by the CC-NB-LRR RPS2 following the cleavage of RIN4 by AvrRpt2, triggering ETI (Axtell and Staskawicz, 2003; Mackey et al., 2003). Figure 3.9 presents the results of this experiment; the control strain *Pto* DC3000 empty vector (EV) induced minimal ion leakage indicating HR was not induced by this strain. After inoculation with *Pto* DC3000(*avrRpt2*), *ham2* mutant lines responded with a similar level of ion leakage to Col-0 at all time-points. This indicates that HAM2 does not function in the mounting of AvrRpt2-induced HR.

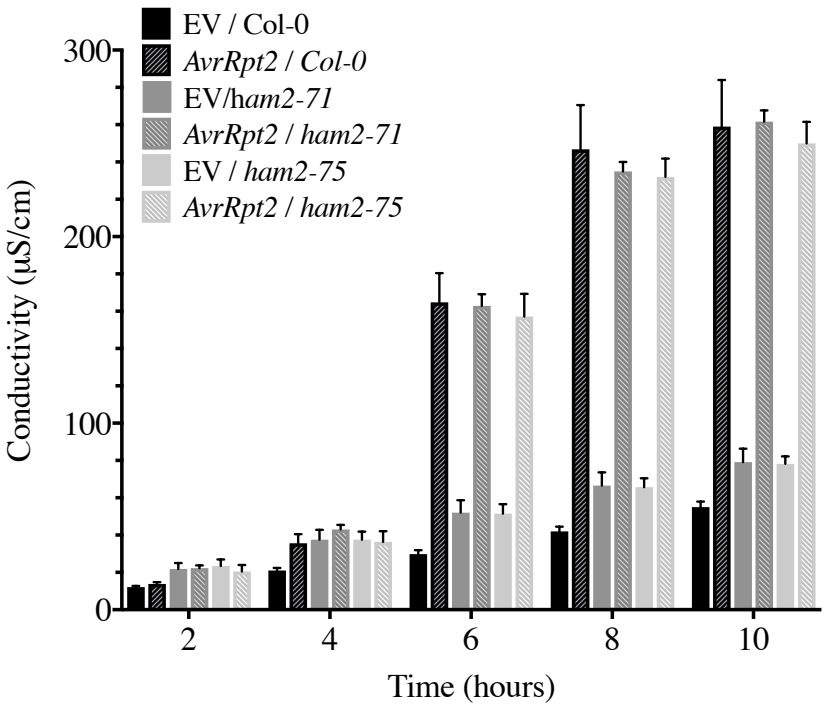


Figure 3.9: AvrRpt2-induced HR is not compromised in *ham2* mutant lines. Leaves of 4-5 week old plants were infiltrated with either *Pto* EV (empty vector) or *Pto(avrRpt2)* at OD_{600}= 0.1. Electrolyte leakage from leaves was measured 2-10 hours every 2 hours after inoculation. No statistically significant difference was observed between Col-0 and *ham2* mutants at each time-point (n=6, two-tailed t-test). Data shown are representative of 2 independent experiments. Error bars indicate standard error.

3.2.4 Defence against *B. cinerea* is not compromised in *ham2* mutants

Given the expression profile of *HAM2* in transcriptomic dataset GSE56094 (Arabidopsis challenged with *B. cinerea*), it was hypothesised that *ham2* mutant lines would have a different phenotype to Col-0 when challenged with *B. cinerea*. Also, published findings have indicated that resistance against biotrophic pathogens can result in enhanced susceptibility to necrotrophs such as *B. cinerea*, and vice versa, due to antagonism of hormonal pathways (Robert-Seilaniantz et al., 2011). To determine if this effect was seen in *ham2-71* and *ham2-75* lines, detached leaf *B. cinerea* infection assays were performed (see Section 2.3.5 for methods description). Figure 3.10 presents the results of this experiment; no difference in *B. cinerea* lesion size was seen between genotypes for all time-points. The trade-off effect of resistance

against biotrophs but susceptibility to necrotrophs was not seen in *ham2* mutants.

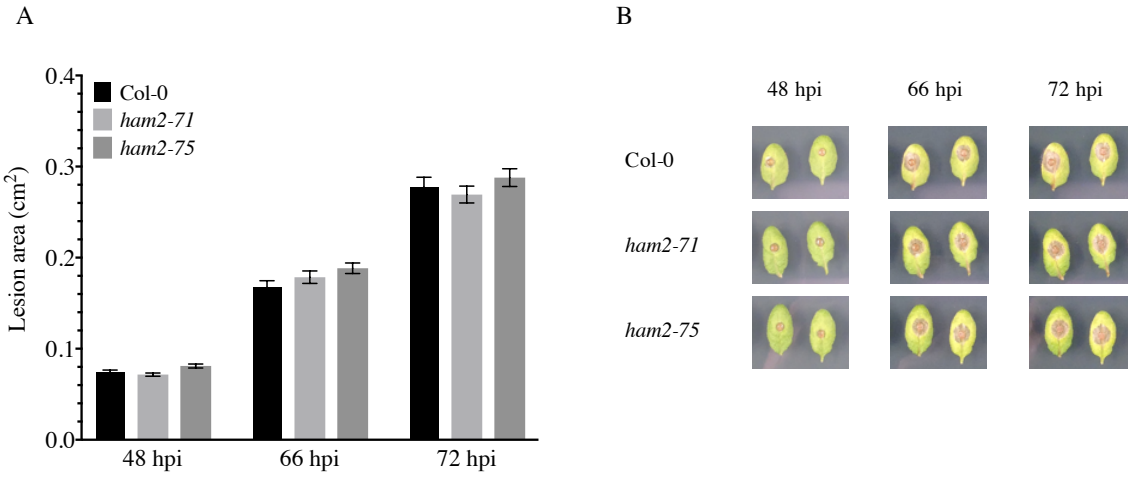


Figure 3.10: Defences against the necrotrophic fungus *B. cinerea* are not compromised in *ham2* mutant lines. (A) Detached leaves of 4-5 week old plants were inoculated with *B. cinerea* var. pepper (1×10^5 spores/mL) and lesion area was measured at 48, 66 and 72 hours post infection (hpi). No statistically significant difference versus Col-0 seen at each time-point ($n \geq 60$, two-tailed t-test). (B) Representative images of *B. cinerea* growth for each genotype tested at each time point. Data shown are representative of 3 independent experiments. Error bars indicate standard error.

3.2.5 Flg22 growth inhibition assay

As described in Section 1.3.5, a plant will either prioritise growth or defence since activation of immunity requires substantial resources. The PAMP flg22 is a 22-amino acid fragment derived from bacterial flagellin which stimulates the FLS2 receptor to elicit PTI (Gómez-Gómez and Boller, 2000). When grown in media containing flg22, seedling growth is inhibited due to constant activation of PTI. The response of *ham2* lines to flg22 was assayed by measuring flg22-induced growth inhibition. Seedlings were grown on 1/2 MS agar plates for 6 days (as described in Section 2.1.1) then transferred to liquid 1/2 MS medium with or without flg22. Seedling fresh weight was measured one week later. As seen in Figure 3.11, the fresh weight of Col-0 seedlings was reduced to an average of 33 mg after exposure to 100 nM flg22 (a weight reduction of 50% compared to 0 nM flg22 conditions). For the *ham2*

mutant lines under 0 nM flg22 conditions, an enhancement of seedling fresh weight can be observed compared to Col-0. Although a similar attenuation of growth can be seen for *ham2-71* and *ham2-75*, the seedlings retain an enhanced fresh weight phenotype after treatment with 100 nM flg22, with a mean fresh weight increase of 15% and 16% versus Col-0 respectively.

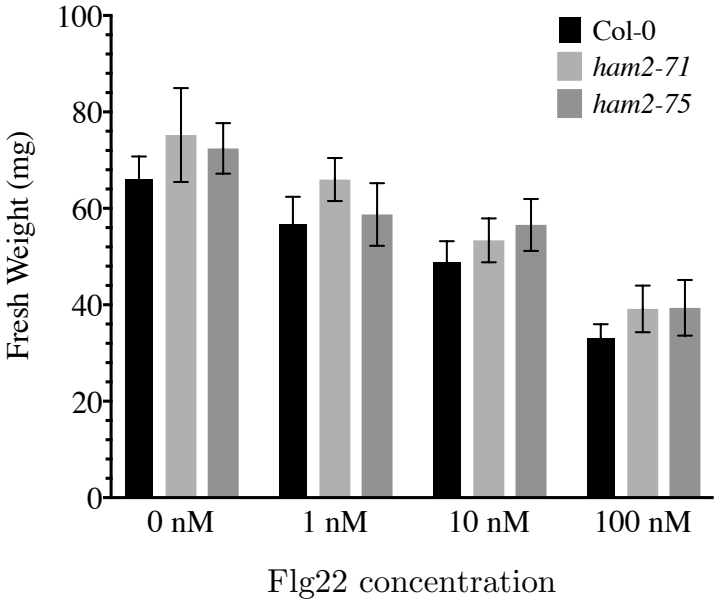


Figure 3.11: *ham2* mutants retain an enhanced fresh weight phenotype under flg22-induced stress. Seedlings were grown on 1/2 MS agar plates for 6 days then transferred into liquid MS with or without flg22 (0 nM, 1 nM, 10 nM or 100 nM). Fresh weight of seedlings was measured one week later. Data shown are representative of 2 independent experiments, n = 6. Error bars indicate standard error.

3.3 HAM2 is a negative regulator of plant growth

3.3.1 Adult *ham2* mutants have enhanced leaf surface area and fresh weight

The finding that *ham2* seedlings have an enhanced fresh weight phenotype (Fig. 3.11) prompted investigation of other indicators of growth and development. Analysis of adult plant leaf surface area showed that *ham2-71* and *ham2-75* lines had an average leaf surface area of 54 mm^2 and 49 mm^2 (27% and 19% greater than Col-0 respectively) (Fig. 3.12B). Furthermore, the average fresh weight was 247 mg for *ham2-71* plants and 219 mg for *ham2-75* plants (32% and 24% greater than Col-0 respectively) (Fig. 3.12C).

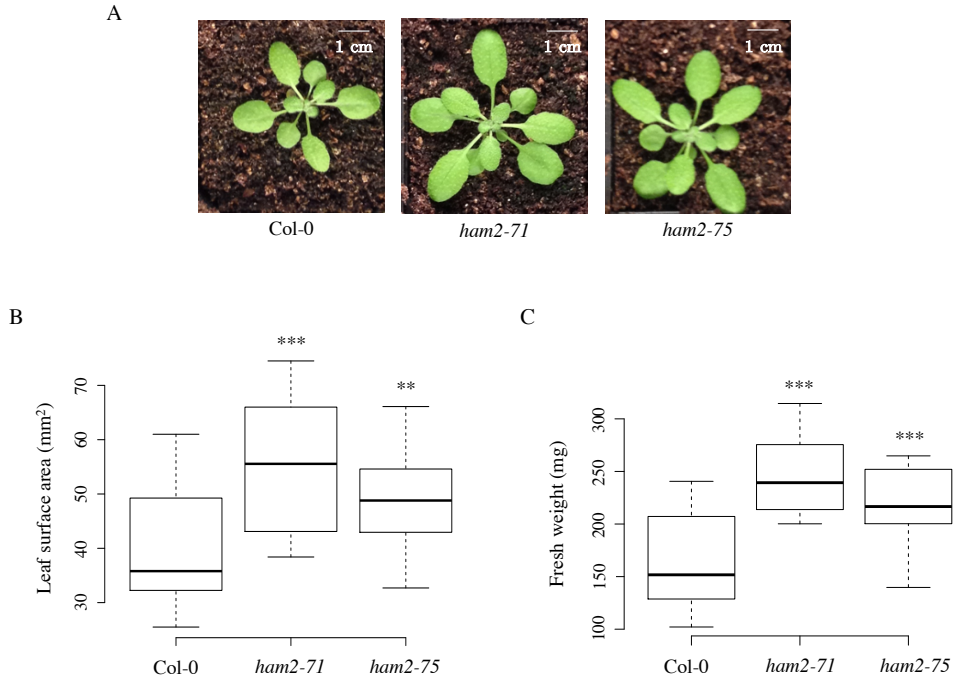


Figure 3.12: *ham2-71* and *ham2-75* mutant lines display an enhanced leaf surface area and fresh weight phenotype. (A) Observed leaf phenotypes, (B) quantified leaf surface area and (C) rosette fresh weight of Col-0, *ham2-71* and *ham2-75* 4 week old adult plants (Section 2.4.1). Data shown are representative of 3 independent experiments. Statistical significance versus Col-0 determined by two-tailed t-test, n $\geq$ 6, ** P $\leq$ 0.01, *** P $\leq$ 0.001. Error bars indicate standard error.

3.3.2 *ham2* seedlings have enhanced root length

To investigate whether enhanced growth was also seen for roots as well as leaves, the phenotype of *ham2* roots at the seedling stage were assessed. Analysis of *ham2-71* and *ham2-75* seedlings showed that primary root length for both *ham2* lines was indeed enhanced: the average primary root length was 72 mm for both *ham2-71* and *ham2-75* lines. Compared to the average primary root length of 65 mm for Col-0, both *ham2* lines display a primary root length increase of 14% (Fig. 3.13B).

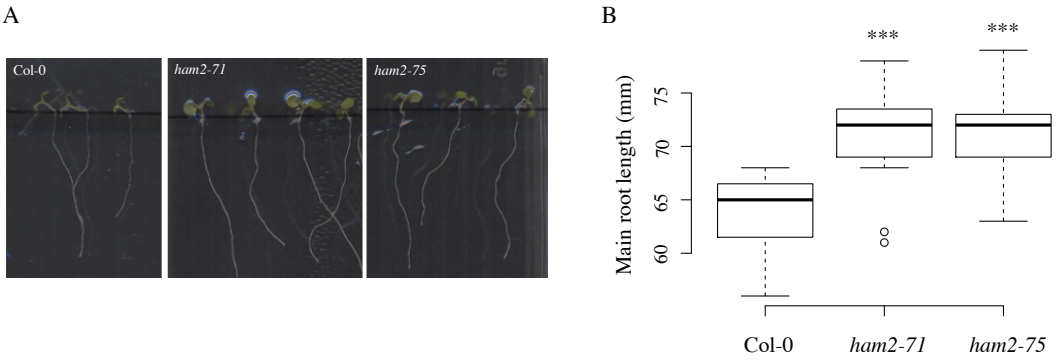


Figure 3.13: *ham2* mutant lines display an enhanced root length phenotype at the seedling stage. (A) Fifteen day old Col-0, *ham2-71* and *ham2-75* seedlings were grown on vertically oriented 1/2 MS agar plates. (B) Quantified primary root length of 15 day old seedlings. Data shown are representative of 2 independent experiments. Statistical significance versus Col-0 determined by two-tailed t-test, n=12, *** $P \leq 0.001$. Error bars indicate standard error.

3.3.3 *ham2* seedlings have enhanced root apical meristem size

The root apical meristem (RAM) is the area of cells beneath the lateral root cap containing stem cells that provide the daughter cells for all cell types in the root. The RAM acts a reservoir of cells for continued root growth and enable the root to extend in length (Sebastian and Lee, 2013). Analysis of root apical meristem size in *ham2* mutants was performed by Anna Gonzalez Gil, a PhD student within the research group. Root tips of 5 day old seedlings were stained with propidium

iodide and imaged with a confocal microscope within 30 minutes of cutting (see Section 2.4.3). The size of the root meristematic zone was determined by counting the number of cells from the quiescent centre (white asterisk in Fig. 3.14A) to the first elongated cell of the transition zone (white arrow in Fig. 3.14A).

As seen in Figure 3.14B, *ham2-71* and *ham2-75* roots have a greater number of cells within the root apical meristem zone compared to Col-0. On average, *ham2-71* seedlings have 48 cells (29% more than Col-0) and *ham2-75* have 51 cells (33% more than Col-0) in the meristem zone.

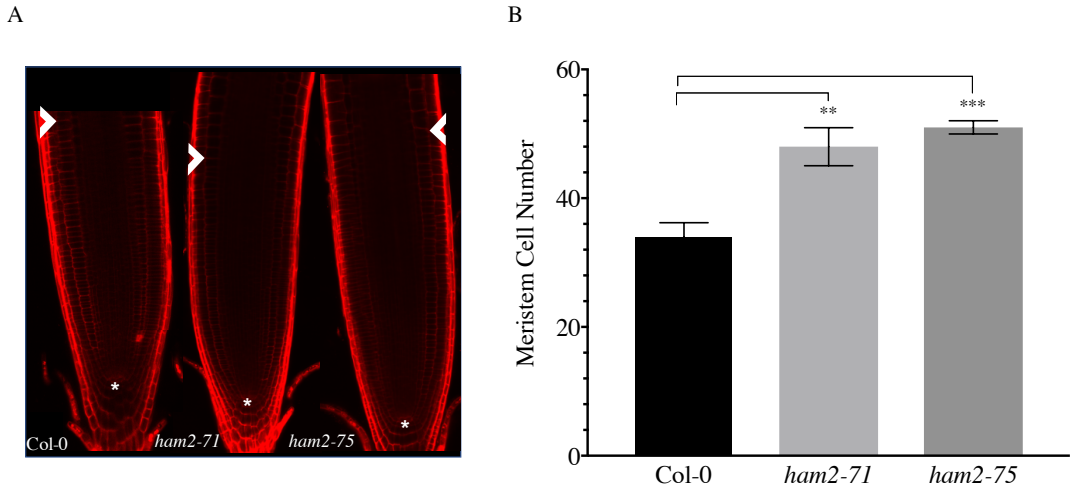


Figure 3.14: Cell number in the root apical meristem is enhanced in *ham2* mutants. (A) Root tips of 5 day old Col-0, *ham2-71* and *ham2-75* seedlings stained with propidium iodide. The meristem zone is indicated between a white asterix (quiescent centre) and white arrow (first elongated cell of the transition zone). (B) Quantification of root meristem cell number in Col-0 and *ham2* mutant lines. Data shown are representative of 2 independent experiments. Statistical significance versus Col-0 determined by two-tailed t-test, n $\geq$ 6, ** P $\leq$ 0.01, *** P $\leq$ 0.001. Error bars indicate standard error.

3.4 Discussion

3.4.1 The function of HAM2 in defence responses

Resistance to *Pto* DC3000

As described in Section 1.3.6, co-ordination of an effective defence response requires huge changes to gene transcription (Jenner and Young, 2005). Since HATs are key transcriptional regulators (Kouzarides, 2007), a reverse genetic screen was conducted to uncover potential roles of HATs in plant immunity. This screen, which was performed previously in the group, found one *hat* (*hag1-6*) to be susceptible and, conversely another *hat* (*ham2-75*) to be more resistant to *Pto* DC3000 (Fig. 3.1). There are no publications describing a role for HAM2 in any aspect of plant immunity to date.

By confirming the *ham2* phenotype in two independent T-DNA insertion lines, it can be concluded that the resistance to *Pto* DC3000 phenotype was due to the specific interruption of *HAM2* expression. It has been published that the remaining *Arabidopsis* MYST HAT, HAM1, acts redundantly with HAM2 through acetylation of H4K5 to regulate gametophyte development and flowering time (Earley et al., 2007; Latrasse et al., 2008; Xiao et al., 2013). Interestingly, work presented in this study demonstrates non-redundant roles for the two MYST HATs, as growth of *Pto* DC3000 was not affected in the *ham1-50* line compared with Col-0.

HAM1 and HAM2 share 87.8% amino acid sequence identity, and the majority of sequence divergence is located at the N-terminus of the protein, close to and within the chromo- (CHRromatin Organisation MOdifier) domain (Fig. 3.15). Chromodomains have been implicated in chromatin remodelling processes through recognition and binding methylated lysines of histone tails followed by recruitment of protein complexes (Fransz and de Jong, 2002). Although the two MYST HATs have been shown to specifically acetylate the same target (H4K5) *in vitro* (Earley et al., 2007), it is possible that they interact with different TFs *in vivo* which could direct them to alternate loci. The ability of HAM1 and HAM2 to bind different TFs may be caused by the structural divergence in the N-terminal region, resulting in the non-redundant roles seen during immune responses. The hypothesis of alternate targeting could be explored experimentally by generating tagged *HAM1* and *HAM2* lines followed by ChIP-seq to identify the genome-wide binding patterns of HAM1 and HAM2 during

infection with *Pto* DC3000.

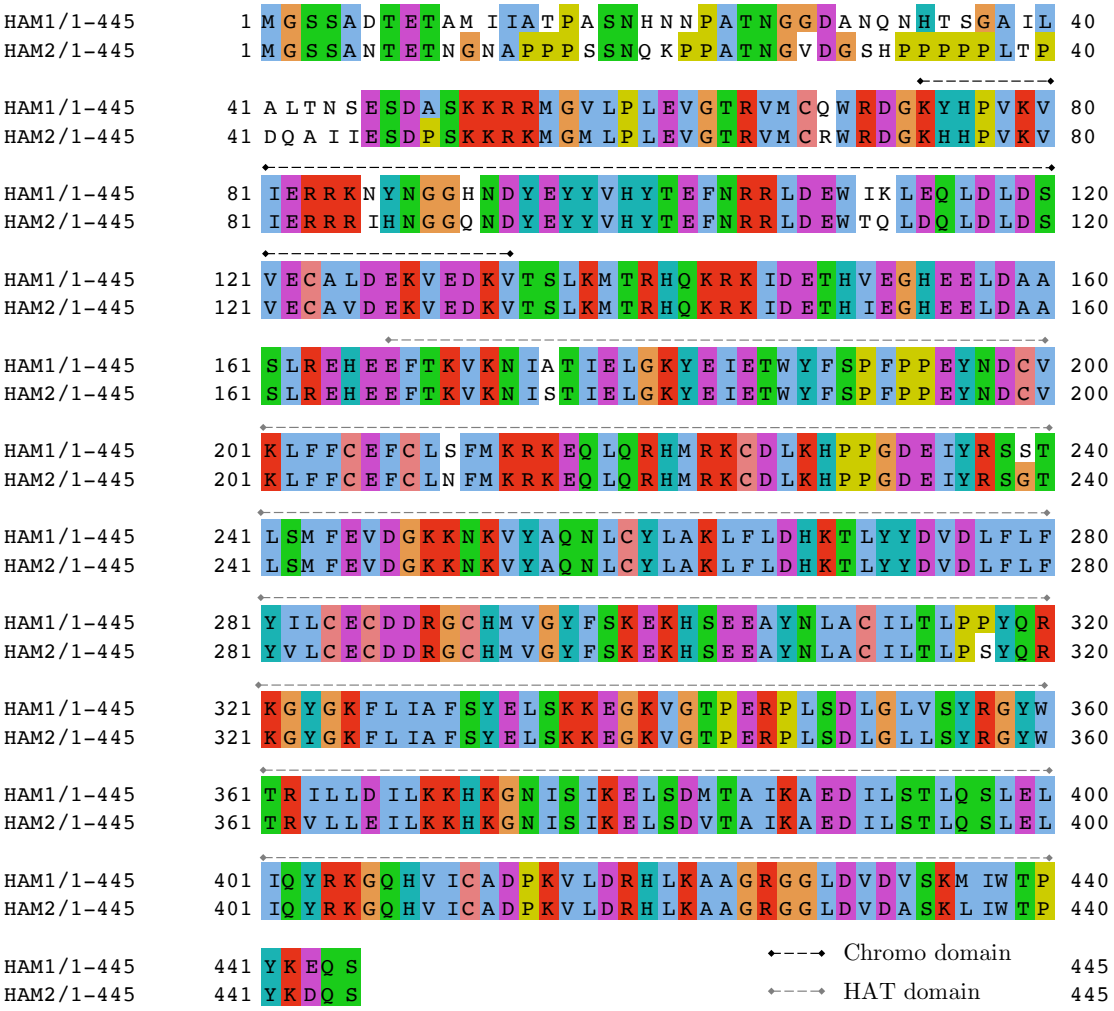


Figure 3.15: Sequence alignment of *Arabidopsis* HAM1 and HAM2. The same colouring for aligned residues indicates conservation of amino acid chemical properties (ClustalX colouring, blue: hydrophobic, red: positively charged, magenta: negative charged, green: polar, cyan: aromatic, pink: cysteine, yellow: proline, orange: glycine). Residues comprising the chromo-domain are indicated with a black dashed line above, residues comprising the HAT domains are indicated with a grey dashed lines above.

Through analysis of *HAM2* expression in published transcriptomic datasets, it was found that *HAM2* was down-regulated following infection with *Pto* DC3000 and even more so following *Pto* DC3000*hrpA*- (Fig. 3.2B). As DC3000*hrpA*- is unable to deliver effectors into the host cell, the difference in the degree of down-regulation

between strains may indicate that an effector is delivered by wild-type *Pto* DC3000 to, directly or indirectly, up-regulate the expression of *HAM2*. This gives weight to the hypothesis that HAM2 is a negative regulator of plant immunity, as sustaining *HAM2* expression levels would benefit pathogen growth through attenuation of defence gene activation. As described in Section 1.5.1, MYST HATs have been implicated in transcriptional repression (Thomas and Voss, 2007). Since findings presented in this chapter indicate that HAM2 is a negative regulator of plant immunity, it could be hypothesised that HAM2 functions to negatively regulate transcription of defence-related genes to moderate the immune response. This could occur through direct transcriptional repression at defence gene loci, or alternatively by transcriptional activation of negative regulators of defence.

The analysis of microarray data performed in this chapter (Section 3.2.2) found that the set of 114 flg22-responsive genes with enhanced expression in *ham2-75* were not enriched for specific GO terms when compared to flg22-responsive genes with equal expression levels between Col-0 and *ham2-75*. This suggests that HAM2 does not repress a specific pathway induced following PTI elicitation. When motif enrichment was performed with the same set of test and background genes, though, several known TF-binding motifs were over-represented. The most statistically significant was the WRKY18 binding motif. WRKY TFs bind to the W-box motif ((T)TGACC/T) in target gene promoters (Eulgem and Somssich, 2007), and are key transcriptional regulators of biotic and abiotic responses in plants (Ren et al., 2010; Hu et al., 2012; Rushton et al., 2010).

WRKY18 is up-regulated >10-fold 30 mins after flg22 treatment (Zipfel et al., 2004) and was recently identified as an important functional hub within a WRKY regulatory network constructed by Choura (2015). Also, constitutive expression of WRKY18 in *Arabidopsis* was found to enhance resistance to *Pto* DC3000 (Xu et al., 2006). Based on these findings and the data presented here, a hypothetical model can be proposed in which, upon infection, *HAM2* is down-regulated causing a "release from suppression" effect of transcription at defence gene loci (such as the WRKY18 binding motif genes), allowing escalation of immune responses. The expression of WRKY18 itself is unchanged in *ham2-75* compared to Col-0 after flg22 treatment, indicating that HAM2 does not directly affect expression of the hub TF WRKY18.

Interestingly, differences in expression after infection with *Pto* DC3000 and *Pto* DC3000*hrpA*- were also observed for WRKY18 binding motif containing genes AT3G26500 and AT3G19010 (Fig. 3.8C and D). The transcriptomic analysis of Col-0 and *ham2-75* lines suggested that these genes are under the repressive action of HAM2 in wild-type plants following flg22 treatment. It follows that, if HAM2 is targeted by an effector to maintain it's repressive function during infection, expression of genes which are under the repressive action of HAM2 would be decreased in *Pto* DC3000 compared to *Pto* DC3000*hrpA*-. As presented in Figure 3.8C and D, this is indeed the case for AT3G26500 and AT3G19010.

For this transcriptomic analysis presented in the chapter, we focussed upon genes that are up-regulated after flg22 elicitation, as these often defence-related genes that contribute to an effective immune response. It would be interesting to also analyse genes that are down-regulated after flg22 elicitation to determine if the resistance phenotype of *ham2-75* could also be due to differential down-regulation post-flg22 perception.

Altogether, to directly correlate HAM2 occupancy, acetylation activity and gene expression before and after flg22 elicitation, the following experiments could be performed. Firstly, HAM2 genomic binding pre- and post-flg22 could be determined by ChIP-seq of tagged *HAM2* lines in a Col-0 background. Next, another ChIP-seq experiment where H4K5ac is pulled down could identify the changes to H4K5ac distribution after flg22 elicitation. Finally, RNA-seq analysis performed before and after flg22 treatment would enable correlations to be made between differential HAM2 binding, H4K5ac levels and differential gene expression.

Response to *B. cinerea*

Since *HAM2* expression is down-regulated in response to *B. cinerea* (Fig. 3.2A) *ham2* mutants were expected to have an alternate phenotype to Col-0 when challenged with the pathogen. Surprisingly, both *ham2* lines responded similarly to Col-0 48, 66 and 72 hours post infection. Sampling for the transcriptomic time course (Windram et al., 2012) ended at 48 hours, and identifying different visible symptoms between genotypes in the *B. cinerea* detached leaf assay before this time-point is difficult. Further experiments at earlier time-points, such as monitoring marker gene expression or phytohormone measurement, would be necessary to further investigate the role of HAM2 in response to *B. cinerea*.

3.4.2 HAM2 is a negative regulator of vegetative growth

Results presented in this chapter also describe HAM2 as a negative regulator of vegetative growth in *Arabidopsis*. At the adult stage, both *ham2* lines displayed enhanced leaf surface area and fresh weight and, at the seedling stage, the primary root length and root apical meristem size was enhanced in both *ham2* lines (Fig. 3.12, Fig. 3.13 and Fig. 3.14). These are novel findings, since other publications assessing *ham* double mutants only describe phenotypes during gametophyte development (Latrasse et al., 2008). In contradiction to results presented here, Latrasse et al. (2008) did not observe abnormal phenotypes of single *ham* mutants, although the authors do not describe the developmental stages and phenotypes assayed for the single mutants. In these studies, it should be noted that Latrasse et al. (2008) used long-day growth conditions, whereas short-day conditions were used in the current study. Also, alternate T-DNA insertion lines were used so differences between phenotypes may arise from differing T-DNA insertion effects. During the current study, double *ham1/ham2* mutants were created with *ham1-50* and the knock-down line *ham2-71* (since double *ham1/ham2* null mutants are not viable (Latrasse et al., 2008)). Even though vegetative growth and immunity phenotypes were not observed in the *ham1* mutant lines, it will be interesting to see if *ham2* phenotypes are exaggerated in the double *ham1/ham2* mutant if there is a measure of redundancy between HAM1 and HAM2 undiscovered in this study.

Links between vegetative growth processes (such as root/shoot meristem growth and leaf morphology) and histone acetylation have been published extensively in the literature. Perhaps the most well-described *Arabidopsis* HAT, the GNAT-family member HAG1, has been shown to be required for cell differentiation promotion, leaf development and root meristem differentiation (Servet et al., 2010). *Arabidopsis* lines with T-DNA insertions in *HAG1*, such as *hag1-6*, show a variety of pleiotropic defects including dwarfism, aberrant meristem function and floral defects affecting fertility (Vlachonasios et al., 2003; Benhamed et al., 2006; Kornet and Scheres, 2009). Here, it is interesting to find that two *Arabidopsis* HATs have seemingly opposing roles in vegetative development which is reflected by their opposing roles in plant immunity. These findings highlight the fact that HATs are diverse in their function, can act as both transcriptional activators and repressors and can have pleiotropic effects on many loci throughout the genome.

3.4.3 Further work

Opportunities for further work have been described throughout this chapter. Other further experiments to be performed in *Arabidopsis* can be summarised in the following points:

1. Create *HAM2* complemented lines to restore the observed phenotypes seen for *ham2* lines
2. Given the difference in *HAM2* expression after infection with *Pto* DC3000 and *Pto* DC3000*hrpA*-, infect the lines with both strains and quantify bacterial growth
3. Perform transcriptomic analysis on *ham2* adult plants infected or not infected with *Pto* DC3000
4. Experiments to explore the relationship between HAM2 and WRKY18, such as ChIP-Seq to determine whether HAM2 directly targets genes containing WRKY18-binding motifs
5. Analyse marker gene expression or phytohormone levels in *ham2* lines following *B. cinerea* infection

3.4.4 HAM2 as a novel target in an agricultural context

The phenotypes of *ham2* mutants described in this chapter present HAM2 as a potential target in an agricultural context. The attenuation of *HAM2* expression in *Arabidopsis* caused enhanced resistance to *Pto* DC3000 with unchanged ETI responses and *B. cinerea* resistance compared to Col-0. Also, the growth and development penalty often associated with active biotic stress responses was less severe in *ham2* lines. Although these beneficial effects are seen from reduced expression of *HAM2* at the seedling and adult stage, negative effects caused by *HAM1/HAM2* double mutation have been presented by Latrasse et al. (2008). It was found that *ham1/ham2* null mutants were not viable and therefore they generated *ham* sesquimutant (*HAM1/ham1; ham2/ham2* and *ham1/ham1; HAM2/ham2*) lines. These plants presented defects in the formation of male and female gametophytes, had reduced silique length and reduced seed numbers in siliques compared to WT. Also, in another study, Campi et al. (2012) found that single *ham1* and *ham2* mutants showed increased DNA damage after UV-B suggesting a role in DNA damage repair.

In light of these studies, it would be ideal if *HAM2* expression could be modulated in a temporally controlled manner, i.e. only in conditions where reduction of *HAM2* expression would be beneficial. A chemical inhibitor targeting the enzymatic activity of HAM2 may offer this opportunity. This approach is particularly appealing, since the knock-down *ham2-71* line displayed similar phenotypes to the knock-out *ham2-75* line, so even partial inhibition of the enzyme could cause the desired effects. In the following chapter, homologues of *Arabidopsis HAM2* were identified in economically important crop species and *in silico* analyses were performed to identify chemical inhibitors targeting HAM2.

Chapter 4

Identification of Inhibitors Targeting HAM2

4.1 Introduction

The findings presented in Chapter 3 indicate that reducing the expression of the MYST histone acetyltransferase HAM2 results in a number of phenotypes in *Arabidopsis* that present HAM2 as a novel target in an agricultural context. The two *ham2* mutant lines tested were found to be resistant to *P. syringae* DC3000, had enhanced leaf surface area at the adult stage and enhanced root length at the seedling stage. Using growth assays following elicitation with the bacterial PAMP flg22, *ham2* seedlings were found to retain the fresh weight phenotype under biotic stress compared to Col-0. Furthermore, resistance to the hemibiotroph *P. syringae* DC3000 did not coincide with increased susceptibility to the necrotroph *Botrytis cinerea* in *ham2* lines.

As previously discussed, we hypothesise that *Arabidopsis* HAM2 (AtHAM2) negatively regulates expression at specific genomic loci to act as a moderator of immune responses. Based on our model without wild-type AtHAM2, defence genes are activated more effectively, with less AtHAM2-induced repressive modifications present at defence loci. This "release-from-suppression" hypothesis explaining the observed phenotypes is currently being pursued by experimental ChIP-seq work within the group.

In this chapter, an *in silico* approach was adopted to identify inhibitors of AtHAM2

catalysis in order to replicate the phenotypes of the *ham2* T-DNA mutants without the use of genetic modification. Previous publications in the literature have described negative effects of disrupting *HAM2* expression, such as increased levels of DNA damage in *ham2* lines following UV-B treatment (Campi et al., 2012). The motivation was to therefore develop a novel agrochemical that would inhibit HAM2 when applied in specific conditions to different agricultural crops if the enzyme retains high cross-species conservation.

4.1.1 Development of inhibitors targeting MYST HATs

Accumulating evidence linking the dysfunction of HAT activities to human pathologies (Yang, 2004; Heery and Fischer, 2007; Barnes, 2009; Johnson et al., 2013) has led to the targeting of the MYST subfamily of HATs for the development of novel therapeutics. Multiple publications have utilised *in silico* and *in vitro* techniques for the identification of MYST inhibitors, leading to the description of candidate inhibitors with the desired activity in cell-based assays. For example, Coffey et al. (2012) conducted a high-throughput *in vitro* screen of 80,000 compounds against the human MYST Tip60 in which an isothiazole (NU9056) was a candidate hit. Subsequent *in vitro* HAT activity assays following treatment with the compound reduced the activity of Tip60 more specifically than other HATs, and dosing a prostate cancer cell line with 24 μM (GI_{50}) NU9056 significantly reduced cellular proliferation.

In another study, Gao et al. (2014) took a rational drug design combined with *in silico* modelling approach to target Tip60, and used the structures of the natural ligand (acetyl-CoA) and a known HAT inhibitor (pentamidine) as scaffolds. Analysis of the electrostatics of the binding pocket guided the variation of different moieties at key positions, and virtual docking led to the identification of a compound (TH1834) that specifically inhibits Tip60 activity *in vitro* and increased apoptosis in breast cancer cell lines (Gao et al., 2014). Though the mechanisms of inhibition were not described in these publications, these studies are encouraging as they indicate that subfamily-specific modulators of MYST activity can be identified through use of *in silico* techniques.

4.1.2 *In silico* docking for inhibitor identification

Cheminformatics and *in silico* docking techniques have become key components of drug discovery programs, allowing researchers to focus time and resources on can-

didates that are more likely to have a desired activity (Sousa et al., 2013). The fundamental aim of *in silico* docking is to predict which small molecules from a library will bind to a target and how strong the interactions will be. Docking screens can either be ligand-based (indirect) or structure-based (direct) (Sliwoski et al., 2014).

In ligand-based screens, molecules that are known to bind the target protein are used to derive pharmacophores; these are models that define the structural characteristics that a molecule must have to bind the target of interest (Wolber and Langer, 2005). Since the natural ligand of AtHAM2, acetyl-CoA, is a common cofactor and no plant-selective MYST inhibitors have been published in the literature, in this study a ligand-based approach could lead to candidate inhibitors lacking specificity.

Structure-based screens utilise target structural data to identify compounds that are predicted to bind with high affinity and specificity. Publications reporting desired *in vivo* effects following structure-based *in silico* docking studies are numerous in the pharmaceutical literature (HTS466284 (Singh et al., 2003); Zanamivir (Varghese, 1999); Nelfinavir (Kaldor et al., 1997)).

In silico docking techniques have also been adopted for the discovery of novel agrochemicals. Recently, Hao et al. (2017) used ligand-based rational design to identify selective protoporphyrinogen oxidase (PPO) inhibitors for the control of broadleaf weeds. The structure of sulfentrazone (a PPO-inhibiting herbicide used worldwide) was adapted by decorating the core scaffold with carboxylic ester groups at specific positions. After *in silico* docking of the modified compounds into crystal structures of *Nicotiana tabacum* PPO and human PPO (using Autodock Vina (Trott and Olson, 2010)), they identified compounds with high potency and selectivity towards *Nicotiana tabacum* PPO. Subsequent *in vivo* testing of a candidate compound showed an increase in potency as a herbicidal agent versus sulfentrazone and decreased toxicity against the mammalian HEK293 cell line. Other examples of agrochemical research utilising *in silico* docking include López-Ramos and Perruccio (2010), Lindell et al. (2009) and Hao et al. (2012).

4.1.3 *In silico* docking programs

The standard outputs from a docking program are a set of favourable conformations of the test compound (calculated by the docking algorithm) which are then ranked according to docking energy scores (based on a scoring function) (Rognan, 2011). Docking experiments in the past were structurally rigid: both the protein and the test compound bonds were inflexible whilst the algorithm would attempt to dock the compound into the protein. Advances in computing have now permitted the inclusion of flexibility within the protein, test compound or even both concurrently. The main limitations of *in silico* docking experiments are the computational power and time available. Within the scope of this study, it was most appropriate to keep the protein structure rigid and the rotatable bonds within test compounds flexible. The docking program AutoDock Vina (Trott and Olson, 2010) was used in this study due to its high citation rate (it was the most highly cited docking algorithm between 2001-2011 (Sousa et al., 2013)) and local expertise with its usage. The performance of the program was validated in Section 4.4.1.

4.1.4 Homology modelling

Why are homology models needed?

A fundamental requirement of a structure-based docking screen is structural data for the target of interest. Multiple approaches can be used to deduce the structure of a protein, for example by NMR spectroscopy or X-ray crystallography. After experimental deduction and quality assessment, structural data is deposited into repositories such as the Protein Data Bank (PDB (Bernstein et al., 1977)). Although there have been major technological advances in structure determination methods, and the ever-growing PDB contains 124,626 structures to date, the methods are experimentally challenging and can cause issues when time and resources are limited. For example, an X-ray crystallography experiment can have issues with expression of the native-state protein at sufficient levels, structural alterations can be caused by purification, and identifying the optimal conditions to obtain high-quality crystals for X-ray diffraction can take from weeks to years.

These challenges have led to a widening gap between the number of structures solved and the number of protein sequences known (Hillisch et al., 2004). In this light, there has been great interest in the prediction of protein structures and several methods have been developed for this purpose. Amongst these methods, homology

modelling is recognised as the most reliable when an experimentally determined template structure is available (Dorn et al., 2014).

The creation of an *Arabidopsis* HAM2 homology model was recently published by Raevsky et al. (2016) whilst the work presented in this chapter was being undertaken. Raevsky et al. (2016) investigated whether the two predicted binding sites of AtHAM2 (one site for Ac-CoA and the other for protein-partner binding) were likely to be functional by studying the AtHAM2 structure. To generate a homology model, they used the SWISS-MODEL server (Arnold et al., 2006) and PDB repository code 2GIV (resolution: 1.94 Å) as a template. Their results indicate that AtHAM2 is likely to bind Ac-CoA in a highly similar manner to the human homologue of AtHAM2, HsKAT8. Although there is no predicted homologue of HsMSL1 (a regulatory protein that binds HsKAT8 (Huang et al., 2012)) in *Arabidopsis* based on sequence searching, they hypothesise the existence of a HsMSL1 functional homologue due to the similar binding pattern formed after *in silico* protein-protein docking of HsMSL1 and the AtHAM2 homology model.

The use of the Raevsky et al. (2016) AtHAM2 model was not appropriate for this work because the model was created using HsKAT8 complexed with Ac-CoA as a template (PDB ID: 2GIV). The Ac-CoA binding pocket of the AtHAM2 homology model was therefore generated in an "open" conformation. It has been reported that presenting the full volume of a binding pocket during a docking experiment may yield false positive results (Uzunova et al., 2016). To minimise these occurrences, a template structure without a ligand co-crystallised in the binding pocket was utilised to create AtHAM2 homology models in the current study.

The basis of homology modelling

Homology modelling (also known as comparative modelling) is based on the observation that similar amino acid sequences form highly similar structures. In fact, Rost (1999) analysed more than one million sequences alignments between structurally determined protein pairs to derive a limit of this rule, the outcome of which is summarised in Figure 4.1. The findings indicate that sequence alignments of $\geq$ 50 residues with $\geq$ 40% sequence identity will almost certainly adopt a highly similar structure. Therefore a protein with an empirically-solved structure can be used as a template to predict an unknown structure (target) if the two sequences share enough homology.

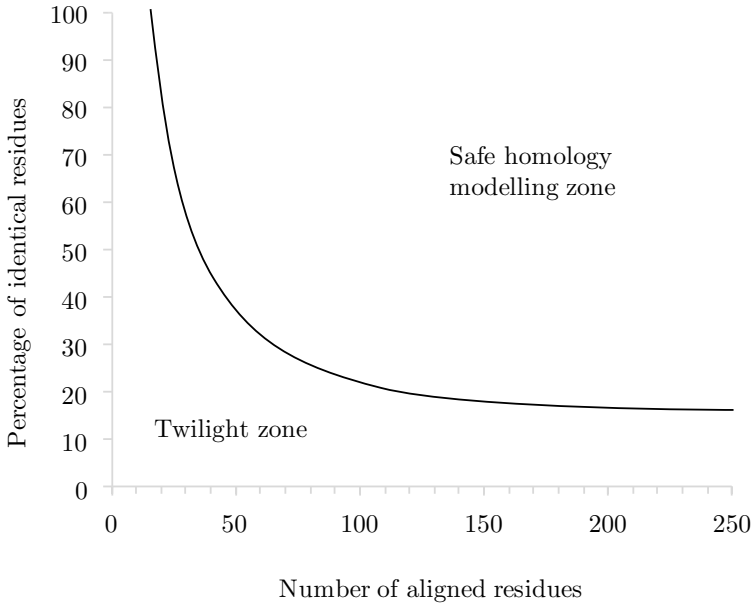


Figure 4.1: The "safe" and "twilight" zones of homology modelling. If the alignment length and percent identity of two sequences lie in the "safe" homology modelling zone, the structures will fold in an almost identical manner. The structure of a protein cannot be predicted reliably if the alignment metrics lie within the "twilight zone". Adapted from Rost (1999).

The process of homology model creation can be summarised in the following steps:

1. Template identification and sequence alignment
2. Generation of the backbone
3. Modelling of loop regions
4. Side-chain generation and optimisation
5. Model optimisation
6. Model quality evaluation

Template identification

Identification of template proteins can be performed by searching the PDB for sequences with similarity to the target sequence using the BLAST or FASTA servers (Altschul et al., 1990; Pearson, 1990). The output includes metrics such as an E-value, percent identity and coverage, but it is of importance to analyse the resolution

of structures and whether ligands/inhibitors were co-crystallised when choosing the optimal template.

Sequence alignment and correction

Once the optimal template has been identified, a target-template sequence alignment must be made. There are many sequence alignment algorithms available that either perform local or global alignments. Global alignments, like the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970), align the query and reference sequence from end-to-end. For local alignments, such as the Smith-Waterman algorithm (Smith and Waterman, 1981), the query is matched with a substring of the reference. In the context of homology modelling, the choice between local and global algorithms depends on coverage of sequence regions. For example, a global alignment would not be suitable when aligning a multi-domain sequence to a homologous single domain: a local alignment between only the two homologous domains would produce a higher score and more biologically meaningful alignment.

Alignment programs also differ by their scoring matrices, speed and sensitivity, so suitability depends on the task at hand. In the context of homology modelling, it is also important to manually analyse and refine the alignment using a knowledge-based approach (i.e. ensuring active site and catalytic residues are aligned).

Homology model generation

The target-template alignment and the PDB structure template file (containing Cartesian coordinates of protein atoms) are used by modelling programs to generate three-dimensional models of the target. Several methods of model generation are available, such as assembly from fragment libraries (Greer, 1981); (Blundell et al., 1987) or the segment-matching approach (Levitt, 1992). Another common method is modelling by satisfaction of spatial restraints, where a set of geometrical criteria are converted to conditional probability density functions which can then directly be used as spatial restraints (Sali and Blundell, 1993).

The spatial restraints are derived from stereochemical restrictions within the protein (such as bond lengths, dihedral angle and bond angle preferences), which in turn are derived from molecular mechanics force fields (based on quantum chemistry calculations). In the case of MODELLER (Sali and Blundell, 1993), the CHARMM22 force

field is currently used (MacKerell Jr et al., 1998). The modelling program executes an optimisation method that attempts to minimise the violations of all spatial restraints to produce a three dimensional model. The optimisation procedure employs the variable target function method (VTFM) which uses conjugate gradients and molecular dynamics with simulated annealing procedures for minimisation (Clore et al., 1986). The energy minimisation process aims to find the stable conformation of the protein, such that the net inter-atomic force on each atom is as close to zero as possible (Webb and Sali, 2017).

MODELLER (Sali and Blundell, 1993) is the most commonly used program that creates models by satisfaction of spatial restraints. Due to its citation rate, high ranking performance (CASP (Critical Assessment of Protein Structure Prediction) experiments (Moult et al., 2003)) and because it is free for academic use, MODELLER was used for homology modelling procedures in this study.

4.1.5 Process of work in this chapter

Figure 4.2 describes the flow of work in this chapter. Firstly, the conservation of AtHAM2 was assessed by searching for homologous proteins in species of interest. Next, homology models were created for *Arabidopsis thaliana* AtHAM2, *Brassica napus* (oilseed rape) BnHAM2 and *Solanum lycopersicum* (tomato) SlHAM2. Data presented in Chapter 5 shows that the dysfunction of the GNAT family HAT HAG1 causes undesirable phenotypes in *Arabidopsis*. Therefore a model was also created for AtHAG1 to identify and eliminate inhibitors that could also inhibit HAG1 function. Since all *Arabidopsis* HATs use acetyl-CoA, a ligand-based screen may identify non-specific inhibitors. A direct structure-based screen was therefore used to increase the likelihood of identifying a HAM2-specific inhibitor.

The compound library was then filtered according to agrochemical likeness based on physicochemical properties. The filtered compounds were computationally docked against the AtHAM2 homology model and the human homologue HsKAT8 (PDB repository code 4DNC). The difference in docking affinity was calculated for each compound, and AtHAM2-selective compounds were identified. In order to compare test compound-protein interactions with natural ligand-protein interactions, a novel application named "Sifter" was developed by our collaborator, Dr Charo I. del Genio (University of Warwick, UK). Using Sifter, we were able to rank candidates

according to their likelihood of being a competitive inhibitor of the binding pocket of the natural ligand, Ac-CoA.

The AtHAM2-selective, top-scoring candidate inhibitors were then docked against BnHAM2, SlHAM2 and AtHAG1 homology models. Compounds that created similar interactions in BnHAM2/SlHAM2 compared with AtHAM2, and those that did not dock into the binding pocket of AtHAG1, were short-listed for *in planta* testing.

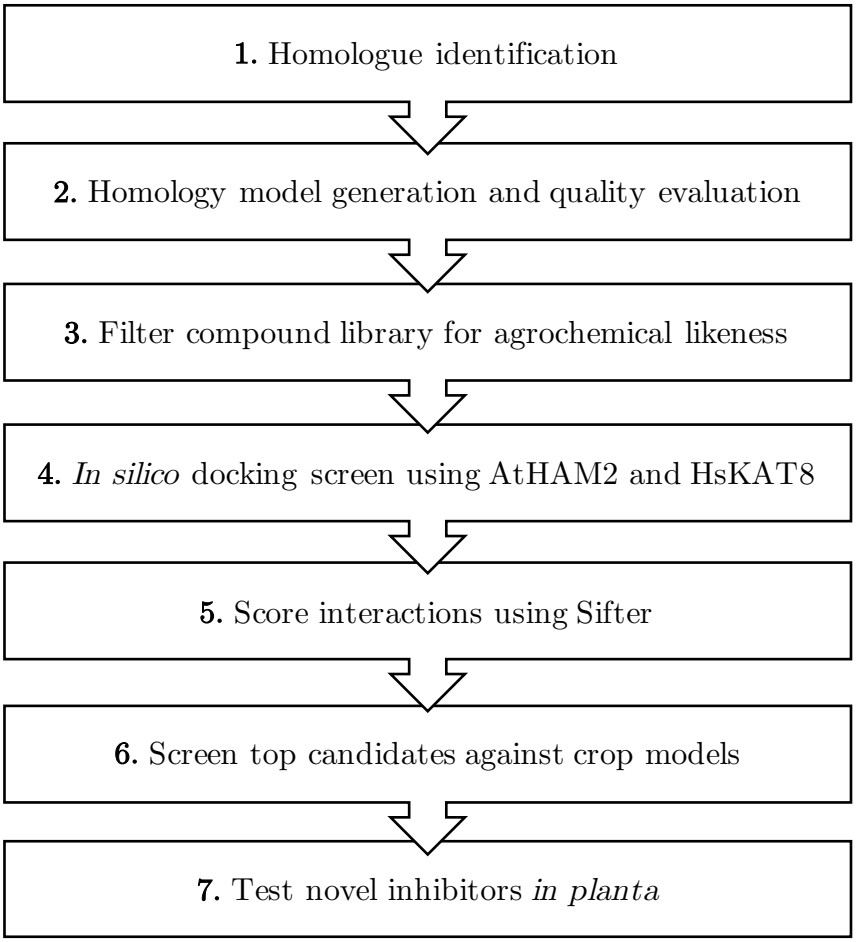


Figure 4.2: Pictorial depiction of the work flow in this chapter, from homologue identification to novel inhibitor discovery.

4.2 Homology modelling of AtHAM2, BnHAM2, SlHAM2 and AtHAG1

4.2.1 Identification of AtHAM2 homologues

To determine the level of conservation across different species, the AtHAM2 amino acid sequence was subjected to protein-protein BLAST searching with the parameters detailed in Section 2.7.1. We were specifically interested in the conservation of HAM2 in *Brassica napus* (oilseed rape) and *Solanum lycopersicum* (tomato) as they are both crop plants of significant agricultural value to the UK and worldwide. World production of tomato is estimated to be 100 million tons fresh fruit from 3.7 million hectares, and is reported as the second most important vegetable crop following potato (of the United Nations, 2001). The level of conservation between *Arabidopsis* HAM2 and human MYST HATs was also evaluated to determine the likelihood of an adverse toxicology profile arising in later stages of safety testing.

The percent identities of homologues identified in these searches are presented in Table 4.1. HAM2 is very well conserved in oilseed rape, with 94.1% sequence identity for its closest homologue of AtHAM2. HAM2 is also well conserved in tomato, with 84.2% identity. Also, a homologue of *Arabidopsis* HAM1 was identified in oilseed rape but was not in tomato. This confirms the findings of Cigliano et al. (2013) who, when investigating histone modifying enzymes in tomato, identified a single HAM protein which is expressed in all organs of the fruit. Although these are predicted genes and activities have not been confirmed *in planta*, it is likely that they produce functional MYST proteins given the high degree of sequence identity with AtHAM2. Other crop species in which HAM2 is highly conserved include *Camelina sativa* (an oilseed crop, 95% identity), *Raphanus sativus* (radish, 93% identity) and *Brassica oleracea* (95% identity). These encouraging results suggest that HAM2 is highly conserved in important crop species and inhibitors developed initially in *Arabidopsis* may also inhibit HAM2 homologues to produce the desirable phenotypes seen in the *ham2* mutants.

Although to a lesser extent compared with plant species, AtHAM2 shares 49.5% sequence identity with the closest human homologue, HsKAT8. This important factor was taken into account during the screening process. Given the degree of difference between HsKAT8 and AtHAM2, and the published success of identification

of a fungal HAT (Rtt109) inhibitor that does not affect the HAT activity of the related mammalian p300 HAT (da Rosa et al., 2013), it is likely that plant-specific inhibitors will be identified amongst the lead compounds.

To compare predicted conserved domains, the amino acid sequences of AtHAM2, BnHAM2, SlHAM2 and HsKAT8 were submitted to the EMBL-EBI tool InterPro (Finn et al., 2016) which collates "signatures" from 14 different databases to identify presence of domains in amino acid sequences. The identified conserved domains and active sites are presented in cartoon form in Figure 4.3: it can be seen that all proteins consist of a N-terminal chromodomain (involved in binding methylated histone-tail lysines (Fransz and de Jong, 2002)) followed by a C-terminal acetyltransferase domain. The predicted catalytic residues are in similar positions within the acetyltransferase domain, suggesting that a common catalytic mechanism may be conserved within these MYST family members.

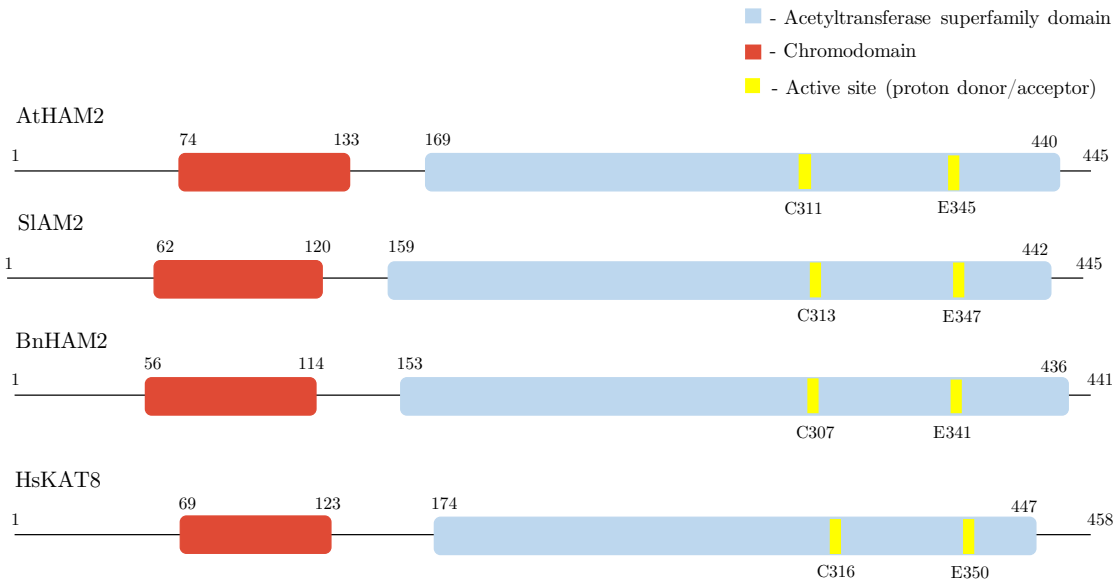


Figure 4.3: The predicted domain structures of HAM2 homologues depicted in cartoon representation. Domain predictions were performed using the EMBL-EBI tool InterPro (Finn et al., 2016).

Table 4.1: Percent identity matrix of HAM2 homologues (NCBI accession numbers in brackets, (O'Leary et al., 2015)). AtHAM2: *Arabidopsis thaliana* HAM2; BnHAM2: *Brassica napus* HAM2; SlHAM2: *Solanum lycopersicum* HAM2; HsKAT8: *Homo sapiens* KAT8.

	AtHAM2 (NP_196536)	**BnHAM2** (XP_013715766)	**SlHAM2** (XP_004250352)	**HsKAT8** (NP_115564)
AtHAM2 (NP_196536)		94.1%	84.2%	49.5%
BnHAM2 (XP_013715766)			85.5%	49.7%
SlHAM2 (XP_004250352)				48.6%
HsKAT8 (NP_115564)				

4.2.2 Template identification

To identify homologues of AtHAM2, BnHAM2, SlHAM2 and AtHAG1 with experimentally determined structures, the amino acid sequence of each protein was subjected to protein-protein BLAST searching entries of the Protein Data Bank (BLOSUM62 substitution matrix, gap penalty=11 and extend penalty=1). As the experimental aim was to identify inhibitors of catalysis, only the catalytic HAT domains of the proteins were submitted to BLAST and subsequently modelled.

The summary statistics of the top 3 "hits" from the PDB for each search are presented in Table 4.2. The E-value (expect) is a BLAST metric that describes the number of hits expected to occur by chance; the value decreases exponentially as the match score increases. From the output data, we were also able to calculate if the template-target alignment exists in the homology modelling safe zone. This was the case for all candidate templates presented in Table 4.2. Extra data were also retrieved from the PDB, such as method of structure determination, level of resolution and co-crystallised ligands.

For all HAM2 proteins, the PDB structure with the highest sequence identity and number of aligned residues was 4DNC (Huang et al., 2012). This is the HAT domain of the human MYST MOF (also known as KAT8) in complex with the MBM

(MOF-binding-motif) domain of MSL1. MSL1 is known to regulate the enzymatic activity of MOF by tethering MOF to a regulatory factor, MSL3. Huang et al. (2012) investigated the structural basis of the regulatory mechanism by solving the complex by X-ray diffraction to a resolution of 2.05Å.

The possibility of MSL1 causing a conformational change in HsKAT8 rendering 4DNC an inappropriate template was assessed by aligning 4DNC with 5J8F (HsKAT8 not co-crystallised with a ligand). The result of the alignment can be seen in Appendix C: the two structures align with RMSD = 0.624 Å between Cα atoms and MSL1 does not cause a conformational change in the backbone structure of 4DNC.

Superimposition of 4DNC (HsKAT8 co-crystallised without Ac-CoA) and 2OU2 (HsKAT8 co-crystallised with Ac-CoA) indicated that a conformational change is induced at α-helix 4 (see Fig. 4.6 for β-sheet and α-helix numbering) as a result of natural ligand binding (see Appendix C). As mentioned in Section 4.1.4, it was decided to create homology models based on unbound template proteins to reduce the occurrence of false positives in the screen. Within the scope of this study, it was not possible to test whether candidate inhibitors would also induce a conformational change similar to the natural ligand.

For the AtHAG1 model, the PDB structure with the highest sequence identity and number of aligned residues was 1YGH (Trievel et al., 1999). This is a high resolution (1.9 Å) crystal structure of the HAT domain of the yeast homologue GCN5 crystallised with glycerol only.

The template proteins chosen for homology modelling are highlighted in bold in Table 4.2.

Table 4.2: Identification of structural templates for homology modelling. (Method: how structure was determined. ResÅ: resolution of solved structure. Ligands co-crystallised: ALY, acetyl-lysine; ZN, zinc; COA, coenzyme A; GOL, glycerol; LYX, N"-(2-coenzyme A)-propanoyl-lysine).

Model	PDB / Chain	E-value	Method	Res. (Å)	Protein	Ligands	Organism	Seq. Identity	No. of aligned residues
AtHAM2									
	4DNC A	3.44E-117	X-ray	2.05	KAT8	ALY, ZN	*H. sapiens*	58.3 %	276
	3QAH A	3.21E-115	X-ray	2.10	" "	ALY, ZN	*H. sapiens*	58.4 %	271
	2PQ8 A	5.45E-114	X-ray	1.45	" "	COA, ZN	*H. sapiens*	58.0 %	271
BnHAM2									
	4DNC A	1.52E-118	X-ray	2.05	KAT8	ALY, ZN	*H. sapiens*	58.6 %	278
	3QAH A	1.75E-116	X-ray	2.10	" "	ALY, ZN	*H. sapiens*	58.7 %	273
	2PQ8 A	2.81E-115	X-ray	1.45	" "	COA, ZN	*H. sapiens*	58.3 %	273
SlHAM2									
	4DNC A	2.64E-118	X-ray	2.05	KAT8	ALY, ZN	*H. sapiens*	57.6 %	277
	3QAH A	6.37E-116	X-ray	2.10	" "	ALY, ZN	*H. sapiens*	58.3 %	273
	2PQ8 A	6.54E-115	X-ray	1.45	" "	COA, ZN	*H. sapiens*	58.0 %	273
AtHAG1									
	1YGH A	1.32E-65	X-ray	1.90	GCN5	GOL	*S. cerevisiae*	57.9 %	159
	5GCN A	6.99E-58	NMR	n/a	" "	COA	*T. thermophila*	55.2 %	159
	1M1D A	2.47E-57	X-ray	2.20	" "	LYX	*T. thermophila*	55.3 %	159

4.2.3 Sequence alignment

The catalytic HAT domains of AtHAM2 (target) and 4DNC (template) were aligned using EMBOSS Water (Smith-Waterman algorithm) (Rice et al., 2000) (parameters can be found in Section 2.7.1). A local alignment program was used to prioritise the alignment of important residues, rather than aligning the sequences end-to-end. The alignment was checked against the 4DNC 3D structure and domain predictions (Fig. 4.3) to ensure residues likely to comprise the binding pocket and catalytic residues were aligned. The resulting alignment is presented in Figure 4.4. The same procedure was carried out for BnHAM2, SlHAM2 and AtHAG1 sequences. These sequence alignments can be found in Appendix C.

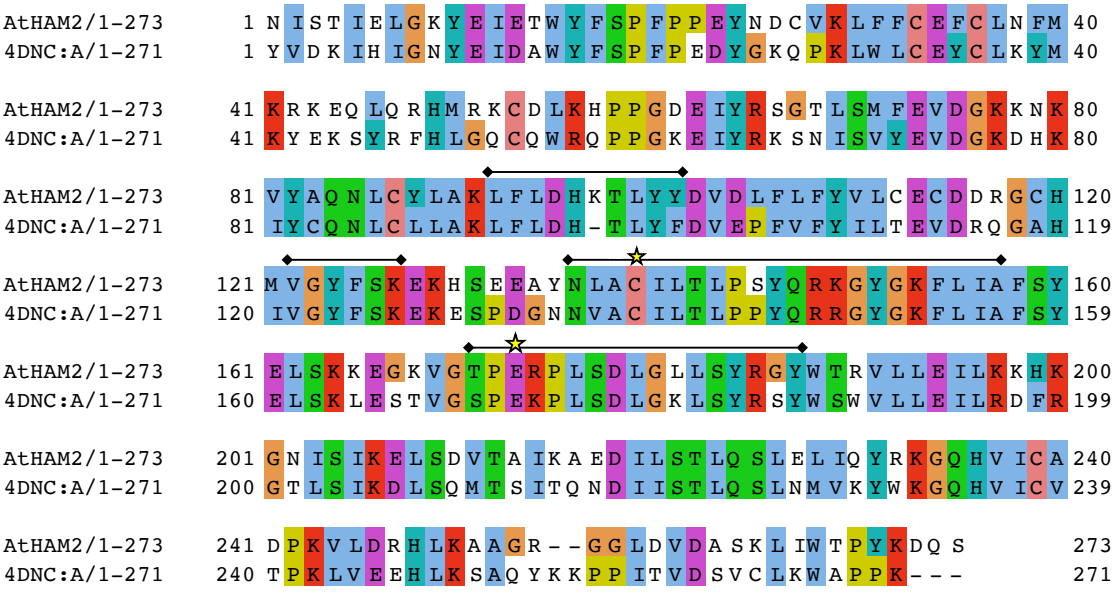


Figure 4.4: Sequence alignment of the HAT domains of HsKAT8 (PDB: 4DNC) and AtHAM2. The same colouring for aligned residues indicates conservation of amino acid chemical properties (ClustalX colouring, blue: hydrophobic, red: positively charged, magenta: negative charged, green: polar, cyan: aromatic, pink: cysteine, yellow: proline, orange: glycine). Catalytic residues are indicated with stars. Residues comprising the binding pocket are indicated with a line above.

4.2.4 Model construction and quality assessment

As detailed in Section 2.7.2, 64 initial homology models were generated using MODELLER (v 9.17) (Sali and Blundell, 1993), followed by 16 loop optimisations yielding 1,024 models in total. All models were ranked by DOPE (Shen and Sali, 2006) and molpdf (Sali and Blundell, 1993) scores which are statistical potentials calculated by MODELLER. Restraint violation profiles for the highest ranking models were assessed, and those with zero heavy violations were identified as candidates. These models were visually assessed to ensure reasonable loop conformations, and the highest ranking model underwent further quality assessment. This procedure was followed for all homology models created in this study.

The results of the additional quality evaluation are presented in Figure 4.5. Firstly, analysis of the DOPE-per-residue scores allowed the detection of unusually high energy regions of the homology model (Fig. 4.5A). Some regions of high energy are seen near residue numbers 50 and 130 when compared to the template. This may be caused by the presence of non-conserved residues (which would adopt different conformations) at these positions compared to the template, as can be seen in the sequence alignment (Fig. 4.4). All of the residues comprising the Ac-CoA binding pocket have similar DOPE scores when compared to the template which confirms the suitability of this homology model for docking studies.

Next, stereochemical quality was assessed by generating Ramachandran plots (Lovell et al., 2003). The performance of the optimisation methods employed by MODELLER was validated here, since the candidate AtHAM2 model had 97.8% residues in the favoured region (98.0% expected), 1.8% in the allowed region (2.0% expected) and 1 residue in the outlier region (Fig. 4.5B). This residue is far away from the Ac-CoA binding pocket and is unlikely to impact *in silico* docking experiments.

ProSA (Protein Structure Analysis) is a commonly used program for the evaluation of protein structure quality. The program makes use of knowledge-based potentials and compares properties of the homology model against the average properties of native proteins (Sippl, 1995). The z-score (Fig. 4.5C), a measure of the overall model quality, compares the total energy deviation against an energy distribution generated from random conformations (Sippl, 1995). Here we can see that the z-score for the AtHAM2 model lies well within the range calculated for experimentally determined proteins. Figure 4.5D, also generated via ProSA, displays the average energies

over 10- or 40-residue fragments of the model (smaller fragments are not suitable for evaluation as they can contain large fluctuations) (Wiederstein and Sippl, 2007). Generally positive values indicate erroneous regions of the model. With a 40-residue smoothing window, all residues have negative energies which confirms the reliability of the model.

These methods of quality assessment were conducted for all homology models, and all chosen models passed this formal evaluation stage (see Appendix C).

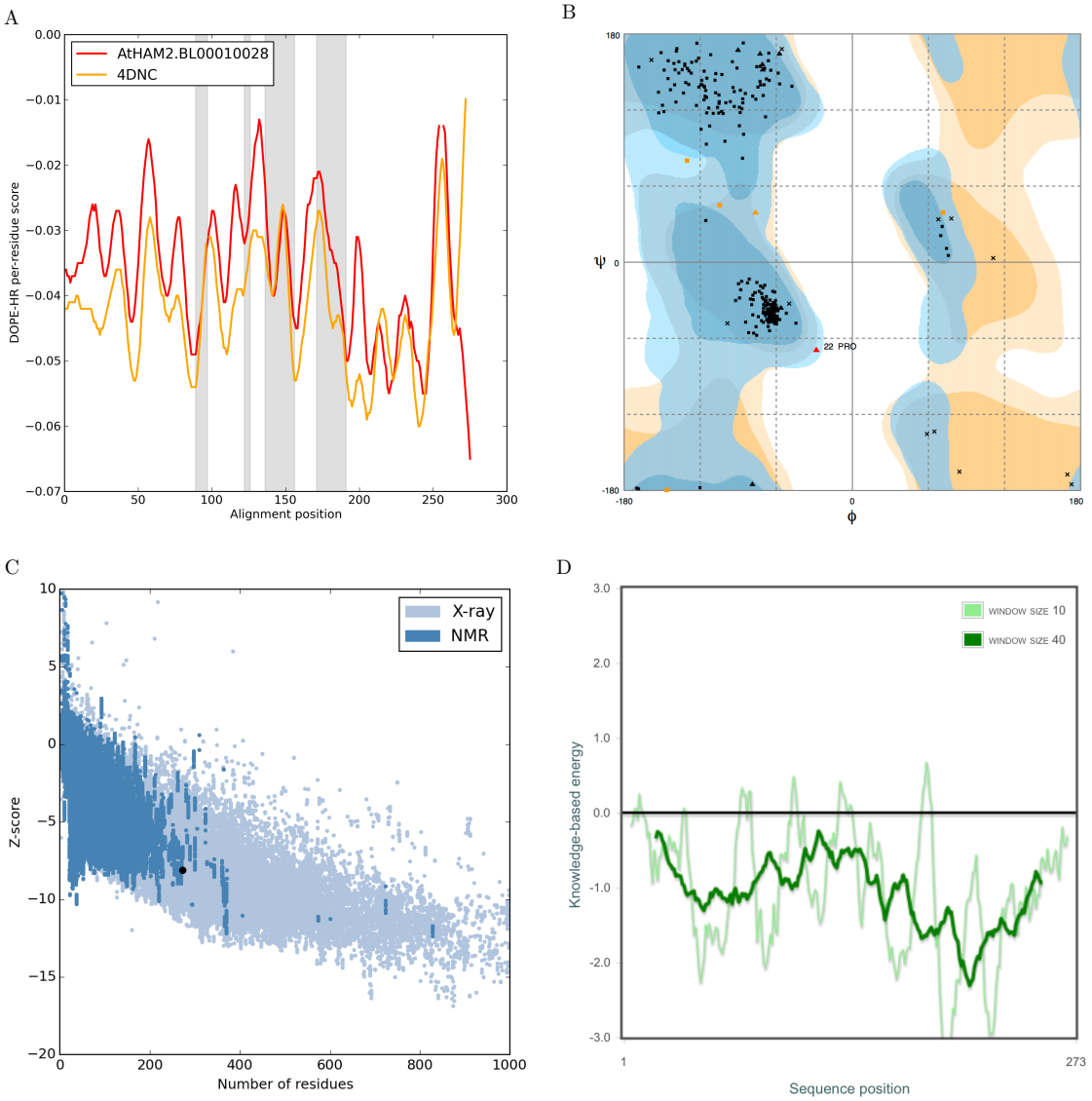


Figure 4.5: Evaluation of AtHAM2 homology model quality. (A) DOPE per-residue score of the template (4DNC, orange) and the target (AtHAM2, red). Residues forming the Ac-CoA binding pocket are shaded in grey. (B) Ramachandran plot analysis. Areas shaded in dark blue, light blue and yellow refer to the "core", "allowed" and "generously allowed" regions respectively. Black dots indicate residue lies within "core" region, orange dots in the "allowed" region and red dots in the "disallowed" region. (C) ProSA z-score of AtHAM2 model (black dot) plotted with Z-scores of all experimentally determined structures currently in PDB (dark blue dots: by NMR, light blue dots: by X-ray). (D) ProSA per-residue interaction energy of the model. Light green line: 10-residue fragment window; dark green line: 40-residue fragment window.

4.2.5 Structural comparison of AtHAM2 and HsKAT8

A structural comparison of the AtHAM2 homology model and HsKAT8 (template) (PDB ID: 4DNC) is presented in Figure 4.6. It is clear that the two structures form a highly similar peptide backbone (RMSD = 0.424 Å between Cα atoms) and the secondary structures formed in HsKAT8 are easily identifiable in AtHAM2. A central core region, which is structurally common to GCN5/PCAF subfamily of HAT proteins (Yan et al., 2000), can also be seen in the AtHAM2 homology model. This core region is made up of three antiparallel β strands (β7-β9) followed by a helix (α3). Another series of β strands and loops (not labelled in Fig. 4.6 for clarity) lead to the C-terminal subdomain. Ac-CoA binds between the central core and C-terminal subdomains and interacts with residues comprising loop-α4 and loop-α3 (Ac-CoA positioning depicted with an arrow in Fig. 4.6).

Although there is contention in the literature regarding the MYST family mechanism of catalysis, it is agreed that a conserved glutamic acid (Glu-350 in HsKAT8) is essential for enzyme activity by acting as a general base for activation of N-ϵ-lysine by deprotonation (Berndsen et al., 2007). The positioning of this residue was analysed in the AtHAM2 homology model and this conserved Glu residue adopts the same conformation in AtHAM2 as in the HsKAT8 crystal structure (Fig. 4.6). This indicates that the MODELLER optimisation protocols were able to find conformations of key residues within conserved regions which match those seen in the HsKAT8 crystal structure.

More structural divergence can be seen in the loop regions, particularly in the loop following β13 and α7. These areas are comprised of residues that are not conserved between HsKAT8 and AtHAM2. A visualisation of conservation between residues is presented in Figure 4.7A and B (guided by the alignment presented in Fig. 4.4). In the central core Ac-CoA binding region, residues are highly conserved which supports the findings of Raevsky et al. (2016) that suggest that AtHAM2 binds Ac-CoA in a similar manner to HsKAT8. Some non-conserved residues are also found close to the "mouth" of the binding pocket (purple region in Fig. 4.6); these will play a key role in identifying plant-specific inhibitors later in this chapter.

Figure 4.7C and D presents the electrostatic surface potential of AtHAM2 and HsKAT8. In both proteins, the "mouth" of the binding pocket is positively charged (blue) whereas the other end of the binding pocket is negatively charged (red). The

negative charge is created by backbone carbonyls as well as acidic residues such as Glu-350. A series of arginines and lysines create a positive charge in the "mouth" region, and these conserved residues form interactions with the negatively charged phosphate groups of Ac-CoA.

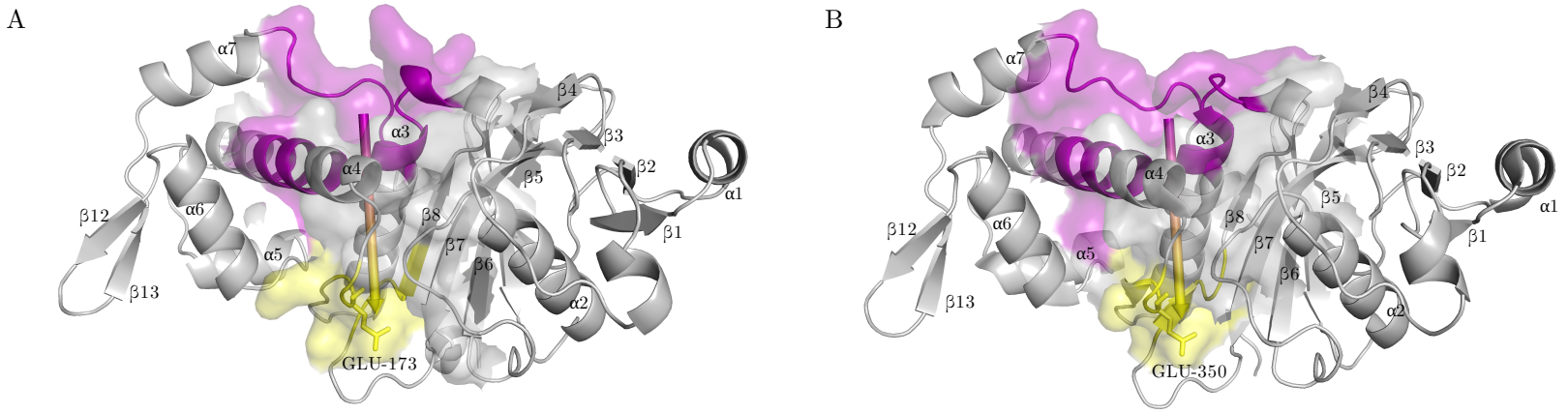


Figure 4.6: Structure comparison of the AtHAM2 model (target) and HsKAT8 (template) (RMSD = 0.424 Å). (A) AtHAM2 homology model and (B) HsKAT8 (template) (PDB ID: 4DNC). α helices and β strands have been labelled according to N-to-C terminal position. Highlighted in purple are the residues that hold the adenosine diphosphate "head" of Ac-CoA during catalysis. In yellow is the "catalytic domain" containing the hypothesised catalytic residue (surface representation of these regions is shown). The arrow indicates the orientation of the long tail of Ac-CoA within the binding pocket (see Fig. 4.8 for HsKAT8/Ac-CoA crystal structure).

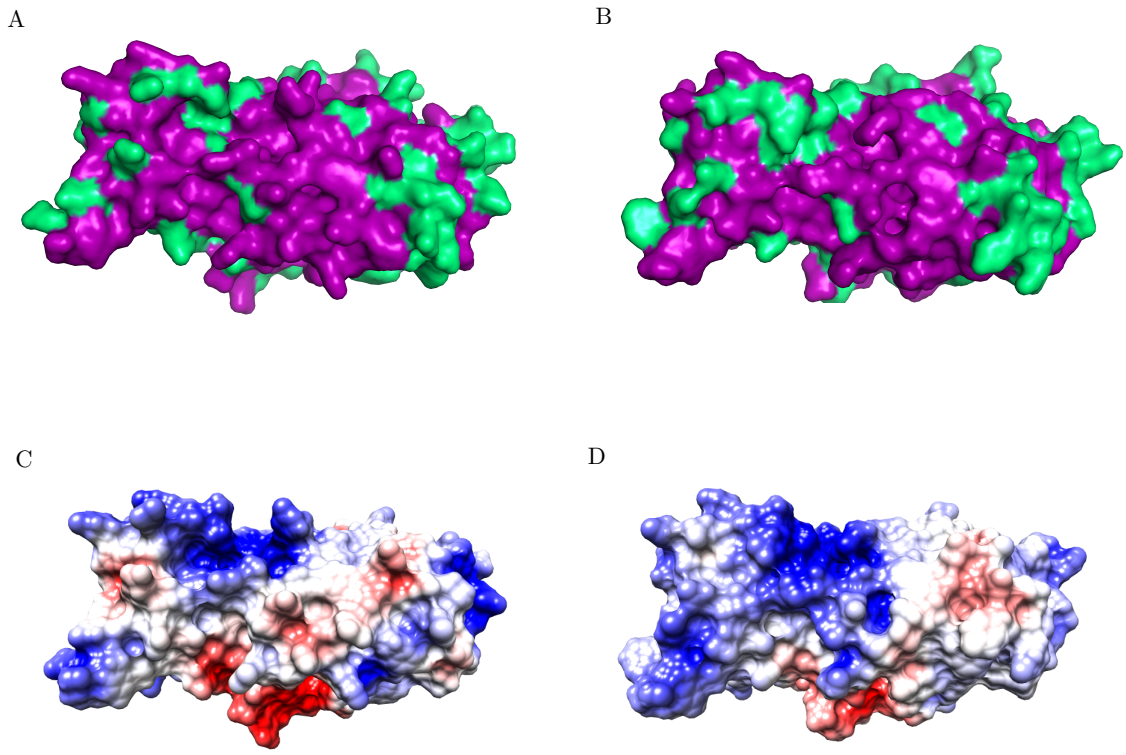


Figure 4.7: Residue and electrostatic surface potential conservation between AtHAM2 and HsKAT8. Purple regions indicate residues with similar physiochemical properties are conserved at that position in both (A) AtHAM2 and (B) HsKAT8 (PDB ID: 4DNC), green regions indicate the presence of dissimilar residues. Electrostatic surface potentials show a similar pattern in (C) AtHAM2 and (D) HsKAT8. Blue indicates positive and red indicates negative charges.

4.3 Compound database preparation and filtering

As described in Section 2.8, two non-overlapping ChemBridge virtual screening libraries were used for the screen. These libraries are comprised of purchasable compounds covering a large chemical space, increasing the likelihood of identifying plant-specific HAM2 inhibitors. The initial number of compounds to screen equalled 1,094,440 unique structures. In order to increase the efficiency of the virtual docking screen, we reduced the number of compounds by filtering according to physiochemical properties of other successful agrochemicals.

Several publications have described quantitative measures of agrochemical likeness by studying the properties of herbicides, insecticides and fungicides currently on the

market (Tice, 2001); (Gandy et al., 2015); (Avram et al., 2014). These physiochemical properties permit the compounds to translocate in the plant vascular system, be foliar-absorbed and pass through cell walls/membranes. Based on the findings presented in these publications, the following criteria were established. Properties of compounds to be screened were within the following limits:

1. Molecular mass $\geq$ 150 and $\leq$ 500 Da
2. Rotatable bonds $\leq$ 12
3. Hydrogen bond acceptors $\geq$ 2 and $\leq$ 12
4. Hydrogen bond donors $\leq$ 3
5. Octanol-water partition coeffecient (Log P) $\leq$ 5

This filtering step excluded 114,768 unfavourable compounds, leaving 979,672 structures to be screened.

4.4 *In silico* docking to identify novel inhibitors

4.4.1 Validation of AutoDock Vina docking algorithm

To validate use of the AutoDock Vina docking algorithm, the protein structure of HsKAT8 co-crystallised with Ac-CoA was used (PDB ID: 2OU2). Ac-CoA was removed from the 2OU2 PDB structure file, the files were prepared for docking and the algorithm was invoked to dock Ac-CoA back into the empty HsKAT8 binding pocket. The best scoring pose is presented in Figure 4.8. The conformation of the "re-docked" Ac-CoA is almost exactly super-imposable when comparing to the original 2OU2 crystal structure from several points of view.

There is a slight variation in conformation at the "tail-end" of Ac-CoA, which may be due to the search exhaustiveness used (this a parameter that controls how many times the algorithm repeats the docking calculations (Forli et al., 2016)). As exhaustiveness is increased, the time taken to perform calculation increases linearly (Jaghoori et al., 2016). Since the conformational differences of the Ac-CoA "tail" are marginal, it was decided to proceed with the same search exhaustiveness used in this docking experiment for efficiency. Overall the results presented here indicate that AutoDock Vina as a suitable docking algorithm to use in this *in silico* screen.

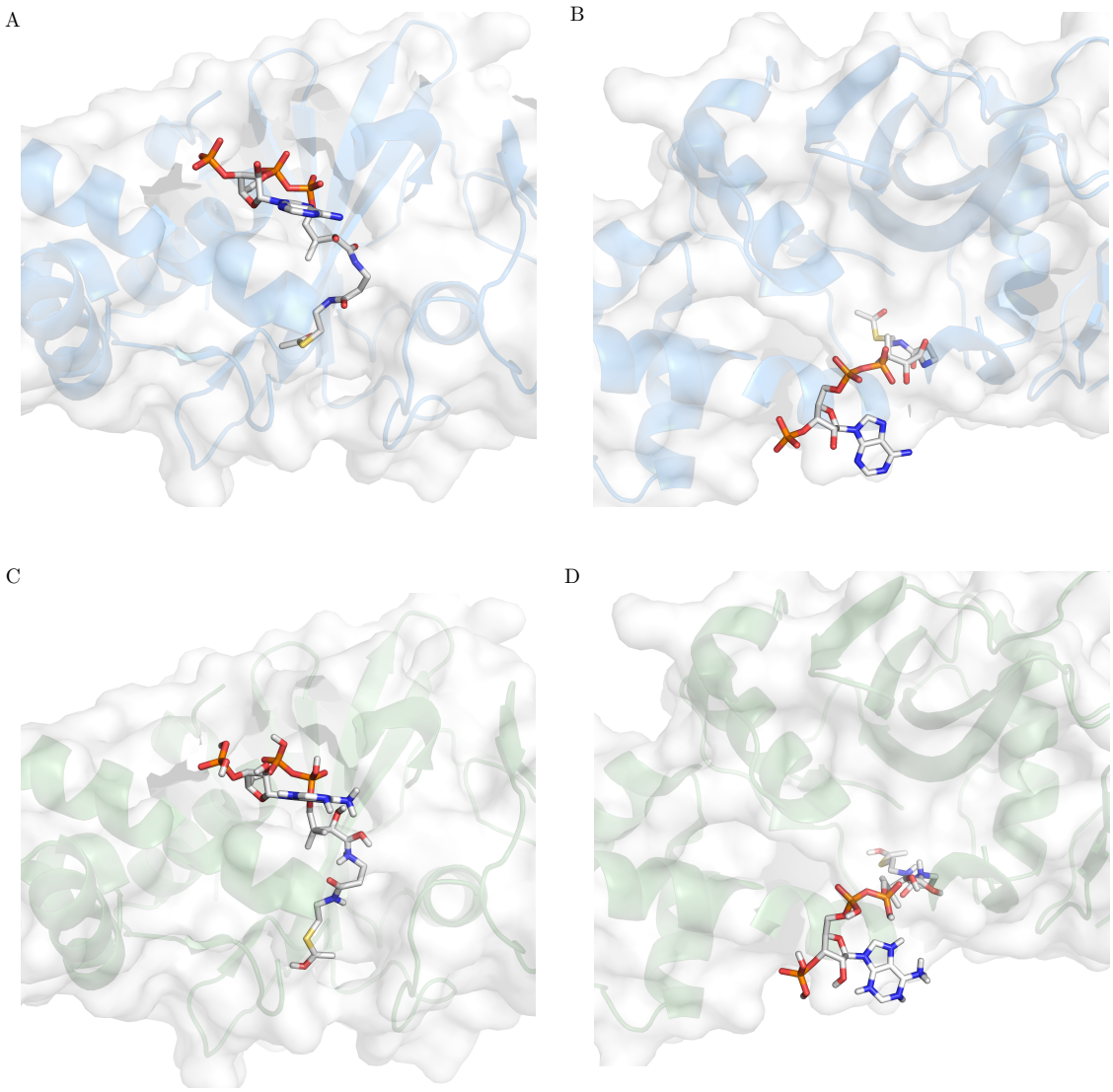


Figure 4.8: Autodock Vina validation: docking of Ac-CoA into HsKAT8. (A) HsKAT8 / Ac-CoA (2OU2) crystal structure (front view). (B) HsKAT8 / Ac-CoA (2OU2) crystal structure (top view). (C) Re-dock of Ac-CoA into empty HsKAT8 binding pocket (front view). (D) Re-dock of Ac-CoA into empty HsKAT8 binding pocket (top view). Protein crystal structures are displayed with cartoon and surface representation. Ac-CoA is displayed in stick representation.

4.4.2 Screen set-up

To identify candidate inhibitors of AtHAM2 that do not dock into the human HAM2 homologue, the library of compounds was screened against the HsKAT8 crystal structure 4DNC and the AtHAM2 homology model in parallel. Before commencing the screen, hydrogen atoms were assigned to the target proteins and library of com-

pounds appropriate for protonation state at pH 7 (reflecting the physiological pH within the plant cell nucleus).

AutoDock Vina also requires a 3D cuboidal grid-box delineated by x, y, z coordinates which defines the search space in which the algorithm attempts to dock the compounds. The grid-box was positioned over the Ac-CoA binding pocket and the same volume of search space was used for all protein structures (26,250 Å) (Fig. 4.9).

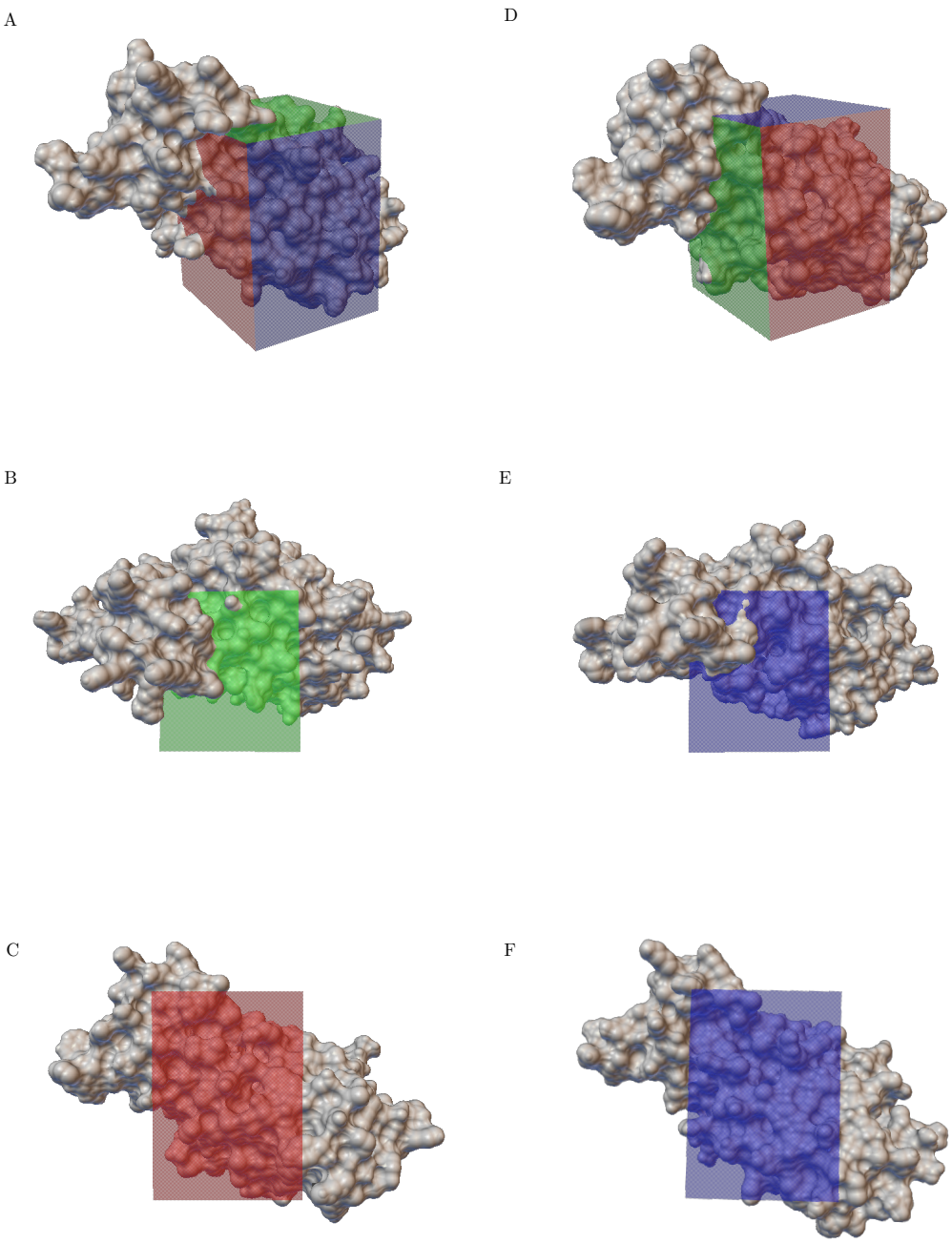


Figure 4.9: Grid-box position over the Ac-CoA binding site of AtHAM2 and 4DNC (HsKAT8). (A) AtHAM2 homology model from a side-view, (B) AtHAM2 from a top view, (C) AtHAM2 from a front view. Blue: x-plane, red: y-plane, green: z-plane. (D) 4DNC (HsKAT8) from a side-view, (E) 4DNC from a top view, (F) 4DNC from a front view. Red: x-plane, green: y-plane, blue: z-plane.

4.4.3 Identifying AtHAM2-selective compounds from screen results

The filtered library of 979,672 compounds was iteratively docked into both AtHAM2 and 4DNC (HsKAT8) within the defined search space. For each compound, the algorithm produced a set of 10 conformations and binding scores for each pose; the highest docking energy score was recorded for each compound. The difference in docking scores between 4DNC and AtHAM2 was calculated for each compound to identify those with high docking scores against AtHAM2 but low scores against 4DNC. An example for the requirement of this step is presented in Figure 4.10. Here, a compound (ChemBridge ID: CB16163267) was identified for its high docking score against AtHAM2 (-8.8 kcal/mol), but the compound also docked with a similarly high score against 4DNC (-8.4 kcal/mol), indicating that this inhibitor would be non-selective for plant species.

The difference in docking energy scores between 4DNC and AtHAM2 for all 979,672 compounds is presented as a histogram and box-plot in Figure 4.11. These scores form a normal distribution with a median of -0.1 kcal/mol, indicating that the compound library has a slight bias of higher docking scores against 4DNC than AtHAM2. To filter out the compounds that have a high docking score with 4DNC, only the 5,791 compounds to the right of the outer fence of the box-plot (Fig. 4.11) were taken forward for further analysis.

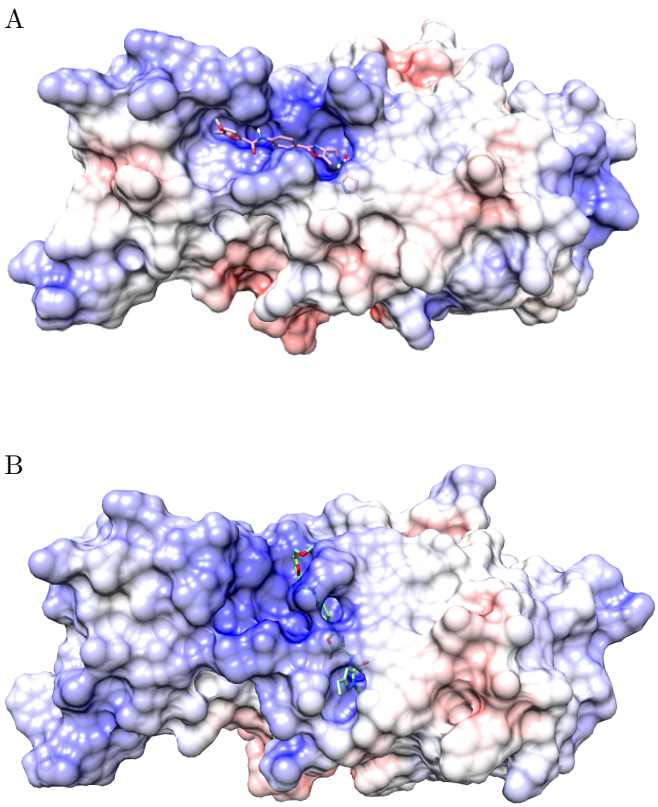


Figure 4.10: A compound that docks with a similarly high docking score into AtHAM2 and HsKAT8. (A) The best scoring pose of CB16163267 docked into AtHAM2 homology model. (B) The best scoring pose of CB16163267 docked into 4DNC (HsKAT8). Both proteins are depicted in surface representation with electrostatic surface potential displayed. The compound is represented in stick form.

4.4.4 Docking AtHAM2-selective compounds against BnHAM2, SlHAM2 and AtHAG1

Since we are interested in the agricultural applications of a HAM2 inhibitor, the 5,791 AtHAM2 selective hit compounds were also docked against homology models of *Brassica napus* BnHAM2 and *Solanum lycopersicum* (tomato) SlHAM2 to identify inhibitors that function across several species. These compounds were also screened against the AtHAG1 homology model because disruption of *HAG1* expression causes undesirable phenotypes in *Arabidopsis*, as presented in Chapter 5.

The same file preparation described for AtHAM2 docking was applied for these pro-

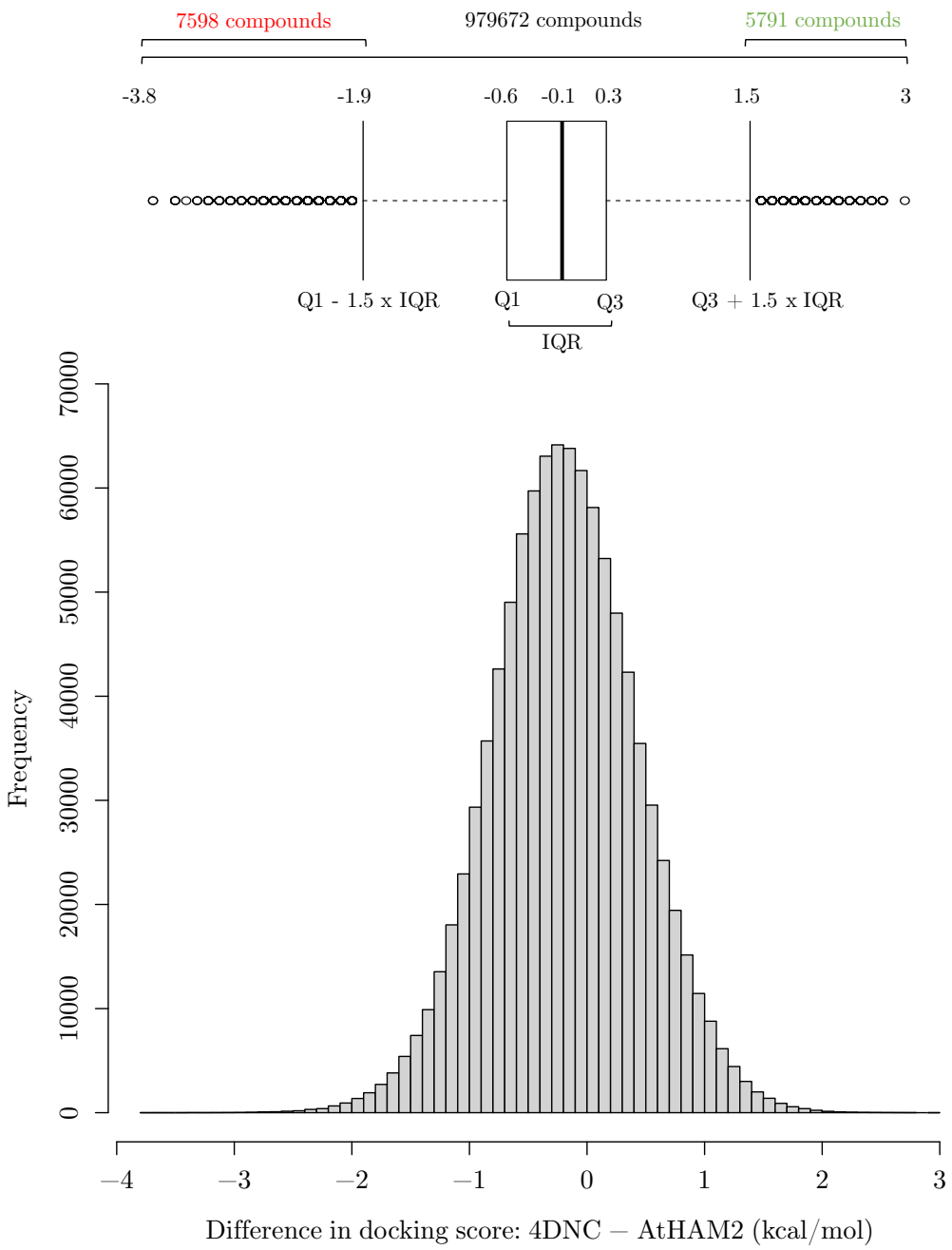


Figure 4.11: Docking scores difference between HsKAT8 (4DNC) and AtHAM2 of all 979,672 compounds screened. IQR: interquartile range, Q1: quartile 1, Q2: quartile 2, Q3: quartile 3, Q4: quartile 4. The same data is presented in box-plot format (top panel) and as a histogram (bottom panel). The 7598 compounds to the left of the left outer fence have a higher docking score with HsKAT8 than AtHAM2; the 5791 compounds to the right of the right outer fence have a higher docking score with AtHAM2 than HsKAT8.

tein models, and the grid-boxes were positioned over the acetyl-CoA binding sites of each enzyme. The highest energy score for each of these docking experiments was recorded and the results are considered in Section 4.4.6.

4.4.5 Comparing Ac-CoA interactions with test compound interactions using a novel program- Sifter

The next stage of candidate inhibitor identification was to compare the poses of test compounds docked into AtHAM2 using the C++ program Sifter. Written by collaborator Dr Charo I. del Genio (University of Warwick, UK), Sifter is a novel computational method that analyses the hydrogen bond formation between a given protein and its natural ligand, and compares them to the hydrogen bond formation between the protein and a test compound. This analysis allows the identification of competitive inhibitors that may preclude binding and entry of Ac-CoA. As input, Sifter requires the protein structure file, the best-scoring docking pose of the natural ligand and each test compound, and the resolution to which volume calculations should be made (in angstroms).
The output metrics of Sifter include the following:

1. Overlap: the fraction of AtHAM2 atoms that interact with Ac-CoA that also interact with the test compound (i.e. a common interaction). A value of one indicates that all AtHAM2 atoms that Ac-CoA interacts with, the test compound interacts with also. A value of zero indicates that the test compound and Ac-CoA do not interact with any of the same AtHAM2 atoms.
2. Overlap dissimilarity: for the AtHAM2 atoms that interact with Ac-CoA and the test compound (overlap), the average excess number of interactions that Ac-CoA forms with respect to the test compound. A negative value indicates that Ac-CoA forms fewer interactions than the compound. A value of zero indicates the same number of interactions are made. A positive value indicates the compound forms fewer interactions than Ac-CoA.
3. RMSD: the root-mean-squared distance deviation between common interactions.

The highest-scoring docking poses of the 5,791 AtHAM2-selective compounds were processed by Sifter and ranked by overlap score. These top-scoring compounds have the highest number of interactions in common with Ac-CoA, so are more likely to behave as competitive inhibitors *in vivo* (the top 10 are presented in Table 4.3; the

top 100 can be found in Appendix C). It can be seen that none of the test compounds have an overlap value of 1, indicating that no compound interacts with all of the AtHAM2 atoms that Ac-CoA interacts with. This factor is redeemed by the presence of negative values in the overlap dissimilarity column; this indicates that the compound makes more interactions with AtHAM2 than Ac-CoA. Compounds with a high-ranking overlap score were taken to the next stage of analysis, with special interest given to compounds with a negative overlap dissimilarity score.

Table 4.3: Top 10 Sifter overlap scorers. The overlap is a value representing the fraction of AtHAM2 atoms that interact with Ac-CoA that also interact with the test compound (i.e. the number of protein atom that interact with both Ac-CoA and the test compound.)

Compound ID	**Overlap**	**Overlap dissim.**	**RMSD**
7_12927	0.7234	-0.4118	0.1632
23_620	0.6809	0.2813	0.1595
46_807	0.6809	0.0625	0.1762
21_18836	0.6596	-0.1613	0.1898
20_10679	0.6383	0.3000	0.1848
9_17598	0.6383	0.3000	0.2029
23_7782	0.6383	0.2000	0.1925
23_10476	0.6383	0.2000	0.1973
40_14139	0.6383	0.2000	0.1770
33_19987	0.6383	0.1333	0.1908

4.4.6 Visual inspection of top-ranking candidate inhibitors

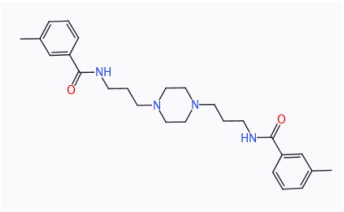
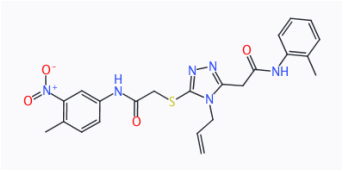
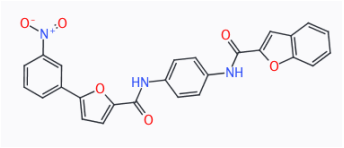
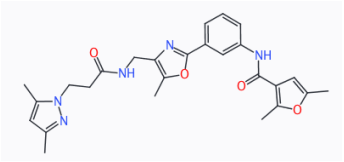
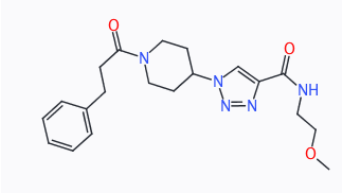
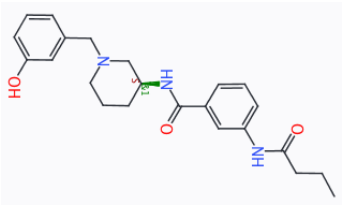
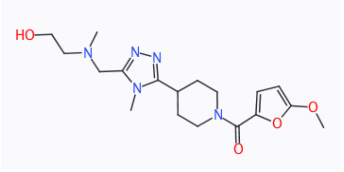
For the next step of analysis, the top 100 ranking overlap scorers previously identified using Sifter were considered. The following criteria were set to select optimal candidate inhibitors from this short-list:

- Inhibitors must interact within the binding pocket of AtHAM2 and with non-conserved residues to further ensure plant-selectivity.

- Inhibitors must dock with a similar orientation with BnHAM2 and SlHAM2 models.
- Inhibitors must not dock within the binding pocket of AtHAG1.

The docking poses of the top 100 ranking compounds against AtHAM2, BnHAM2, SlHAM2 and AtHAG1 were visually inspected in 3D using PyMOL (Schrödinger, LLC, 2015) and in 2D using LigPlot ligand interaction diagrams (Laskowski and Swindells, 2011). From the 100 candidates, 56 docked within the Ac-CoA binding pocket of the AtHAG1 model and were therefore rejected. This result is perhaps not surprising given the structural conservation of the core Ac-CoA binding subdomain between MYST and GCN5/PCAF HAT proteins (Yan et al., 2000). Thirty-seven compounds from the short-list were found to not dock into BnHAM2 and SlHAM2 with a similar docking score and pose compared to AtHAM2 and were also rejected. From the 100 visually analysed, 7 compounds adhered to the criteria above. These compounds are presented in Table 4.4.

Table 4.4: Structure of the candidate lead compounds. These compounds docked with a higher docking score in AtHAM2 than HsKAT8, interacted with non-conserved residues in AtHAM2, did not dock into the binding pocket of AtHAG1 and docked with similar orientation with BnHAM2 and SlHAM2.

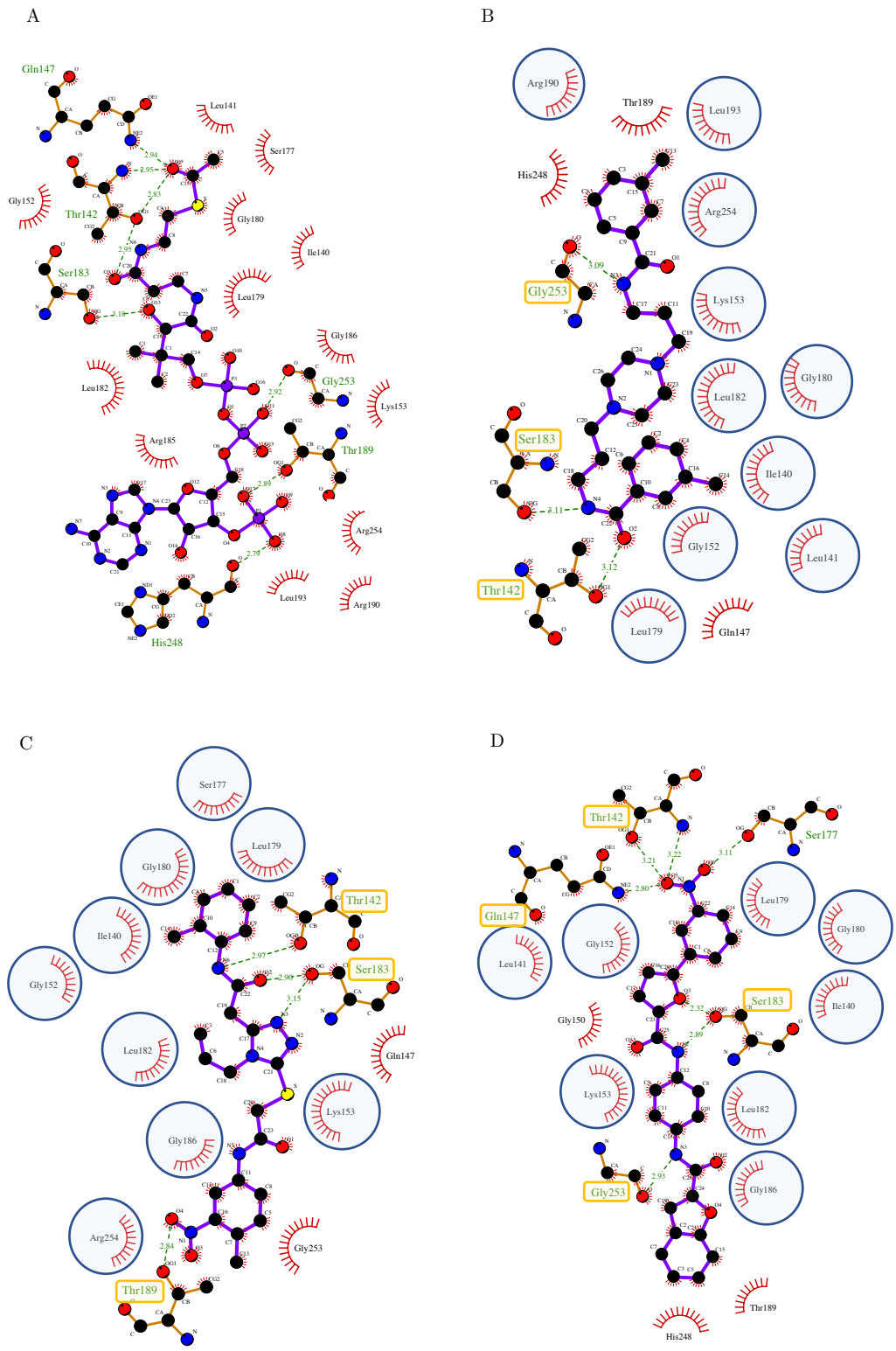
Compound ID	Structure
CB7304179	
CB7919054	
CB7967263	
CB18262368	
CB22474227	
CB26389098	
CB82136042	

4.4.7 Interactions of candidate inhibitors with AtHAM2 binding pocket

Presented in Figure 4.12 as 2D interaction diagrams are the 7 candidate lead compounds, as well as Ac-CoA, docked into AtHAM2. In these diagrams, hydrogen bonds are depicted with green lines (with bond length in Å) and hydrophobic interactions are depicted with red "eyelashes". The molecular interactions made between Ac-CoA and AtHAM2 can be seen in Figure 4.12A and was included in order to assess similarities between Ac-CoA and candidate inhibitor functional groups.

Hydrophobic interactions can be observed between Ac-CoA and 13 neighbouring residues (mainly glycine, leucine and isoleucines). The same residues that create hydrophobic interactions with Ac-CoA often also create interactions with the candidate inhibitors; in Figure 4.12, residues circled in blue indicate that the hydrophobic interaction is common to the compound and Ac-CoA. Candidate inhibitor CB26389098 (Fig. 4.12G) has four hydrophobic interactions not observed between AtHAM2/Ac-CoA, suggesting this compound would make an effective inhibitor.

Eight intermolecular hydrogen bonds can be observed between Ac-CoA and AtHAM2. Three are made between the oxygen atoms of the phosphate groups of adenosine diphosphate "head" of Ac-CoA with backbone carbonyl groups of glycine-253 and histadine-248, as well the hydroxyl group of threonine-189. A further 4 hydrogen bonds are made with the "tail" of Ac-CoA, with the acetyl group forming two bonds with the hydroxyl group of threonine-142 and the carbonyl groups of glycine-147. Although no compound forms eight hydrogen bonds (as Ac-CoA does), one candidate (CB7967263) (Fig. 4.12D) makes 7 hydrogen bonds with AtHAM2, with a terminal nitrogen dioxide moiety mimicking the acetyl group of Ac-CoA. Importantly, all candidate compounds were identified as hydrogen bonding with a residue not conserved between AtHAM2 and HsKAT8 (either the backbone carbonyl of glycine-253 or the hydroxyl group of threonine-189). This increases the likelihood that these compounds will show specific inhibitory activity towards AtHAM2.


A
Gln147
Leu141
Ser177
Gly152
Thr142
Gly180
Ile140
Ser183
Leu179
Gly186
Leu182
Gly253
Lys153
Arg185
Thr189
Arg254
Leu193
Arg190
His248
B
Arg190
Thr189
Leu193
His248
Arg254
Gly253
Lys153
Leu182
Gly180
Ser183
Ile140
Gly152
Leu141
Thr142
Leu179
Gln147
C
Ser177
Leu179
Gly180
Thr142
Ile140
Ser183
Gly152
Leu182
Gln147
Lys153
Gly186
Arg254
Gly253
Thr189
D
Thr142
Ser177
Gln147
Leu179
Gly180
Gly152
Leu141
Ser183
Ile140
Gly150
Lys153
Leu182
Gly253
Gly186
Thr189
His248

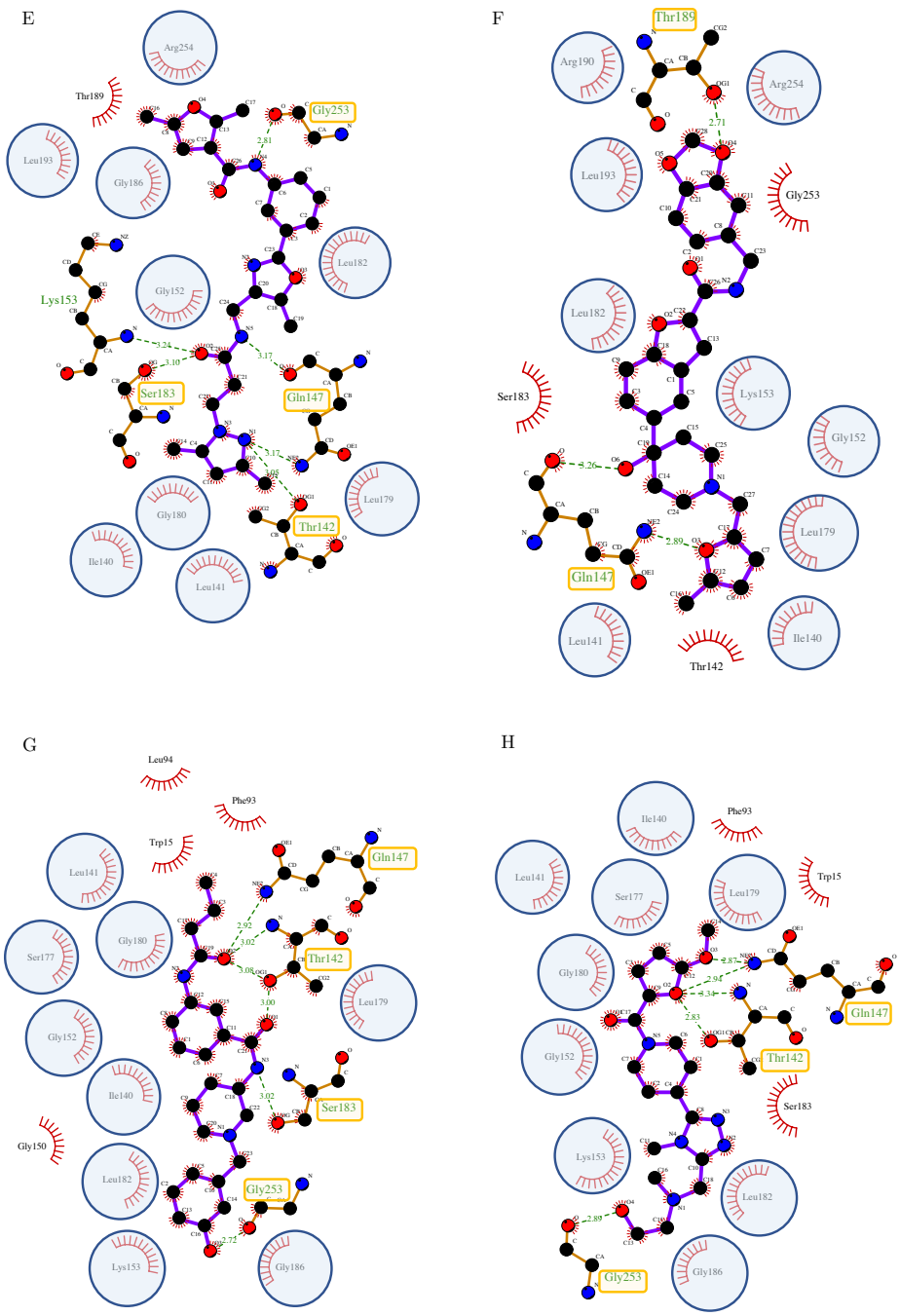


Figure 4.12: 2D interaction plots between AtHAM2 and candidate lead compounds. (A) Ac-CoA, (B) CB7304179, (C) CB7919054, (D) CB7967263, (E) CB18262368, (F) CB22474227, (G) CB26389098, (H) CB82136042. Compounds are represented in ball and stick form, residues are represented by name. Hydrogen bonds are shown as green lines and hydrophobic interactions are indicated with red eyelashes. If residue is circled in blue, Ac-CoA also has hydrophobic interaction with that residue. If residue is highlighted in yellow, Ac-CoA also hydrogen bonds with that residue.

4.5 Initial *in planta* testing

In order to perform initial small-scale tests on the compounds, we utilised the enhanced root apical meristem size phenotype seen for *ham2* seedlings (Section 3.14). It was reasoned that if a compound was applied to Col-0 and a similar enhanced meristem zone size was seen, this change in morphology could be due to loss of AtHAM2 functionality. Alongside Col-0, *ham2-71* and *ham2-75* seedlings were also tested; in this way we were able to determine if any changes were AtHAM2-dependent. To establish the assay, a published compound (NU9056) known to inhibit the human MYST HAT Tip60 was utilised. As described in Section 4.1.1, this compound was identified for reducing the catalytic activity of Tip60 more specifically than other HAT families tested (Coffey et al., 2012) and was therefore chosen to test the suitability of this assay as a potential MYST-specific inhibitor. The following experimental work was performed by Anna Gonzalez Gil, a PhD student within the research group.

The assay was performed as described in Chapter 3 (Section 3.3.3) with the exception that, at 4 days old, the seedlings were treated with control (DMSO) or 2 μM NU9056 for 24 hours. This concentration was chosen as an initial guide as Coffey et al. (2012) published that this was the compounds IC_{50} against Tip60 (determined by *in vitro* HAT activity assay). After treatment, root tips of 5 day old seedlings were stained with propidium iodide and imaged with a confocal microscope within 30 minutes of cutting (see Section 2.4.3). The size of the root meristematic zone was determined by counting the number of cells from the quiescent centre (white asterisk in Fig. 4.13A) to the first elongated cell of the transition zone (white arrow in Fig. 4.13A).

Consistent with previous findings, the root apical meristem zone of *ham2-71* and *ham2-75* showed enhanced size compared to Col-0 (Fig. 4.13). Also, roots of Col-0 seedlings appear to have larger apical meristem zone when treated with 2 μM NU9056 for 24 hours. When *ham2-75* seedling roots were treated with 2 μM NU9056, an increased meristem zone size was seen suggesting that NU9056 may be interfering with a factor other than AtHAM2 to also enhance meristem zone size. Overall, these are encouraging results that indicate that assaying changes to the root apical meristem zone of Col-0, *ham2-71* and *ham2-75* seedlings following treatment with candidate inhibitors could lead to the successful identification of

AtHAM2-specific inhibitors. This will allow the progressing of compounds that are more likely to be successful AtHAM2 inhibitors to adult *in planta* testing.

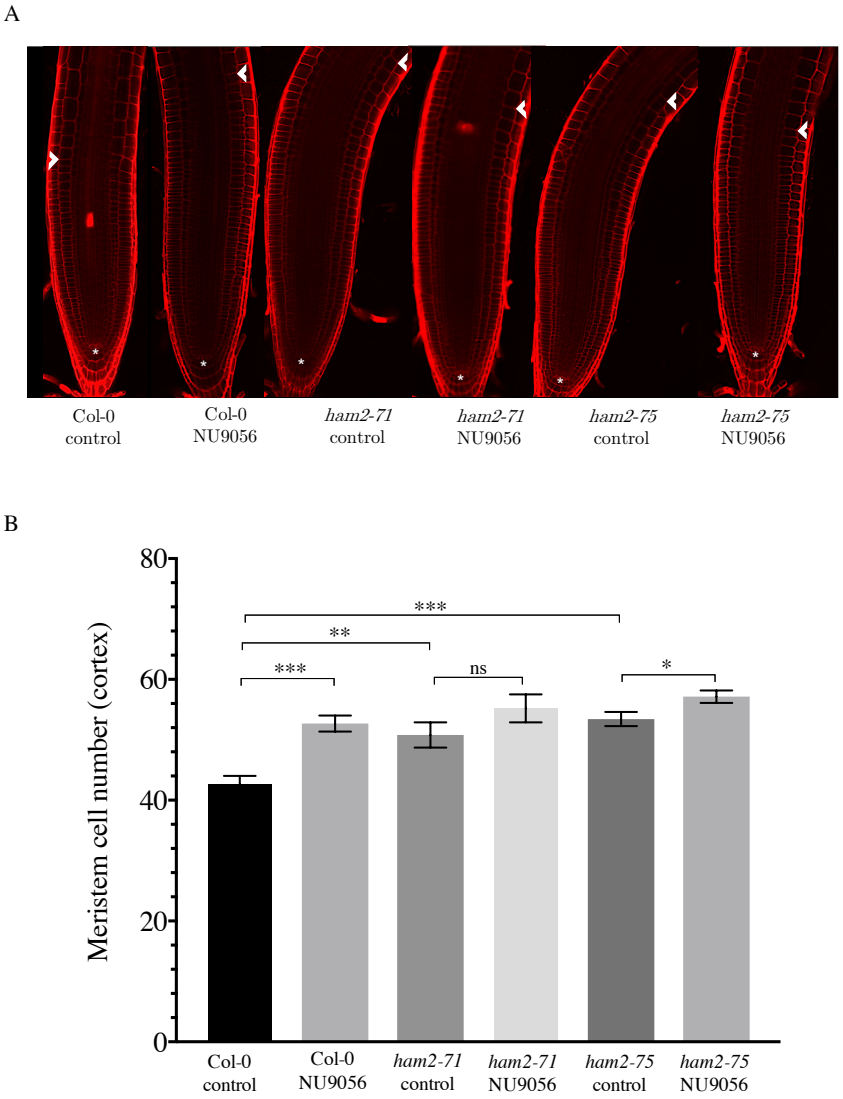


Figure 4.13: Cell number in the root apical meristem is enhanced in Col-0 when treated with NU9056. (A) Root tips of 5 day old Col-0, *ham2-71* and *ham2-75* seedlings treated or untreated with 2 μM NU9056 followed by staining with propidium iodide. The meristem zone is indicated between a white asterix (quiescent centre) and white arrow (first elongated cell of the transition zone). (B) Quantification of root meristem cell number in Col-0 and *ham2* mutant lines. Data shown are representative of 2 independent experiments. Statistical significance determined by two-tailed t-test, n $\geq$ 6, ** P $\leq$ 0.01, *** P $\leq$ 0.001. Error bars indicate standard error.

4.6 Discussion

In the previous chapter, phenotypes such as enhanced adult leaf surface area, resistance to *Pto* DC3000 (without increased susceptibility to the necrotroph *Botrytis cinerea*) and enhanced seedling root length (which is maintained under biotic stress conditions) were identified in *Arabidopsis ham2* mutant lines. These traits are very appealing in an agricultural context because methods used to confer resistance to control crop diseases usually carry yield penalties (Brown, 2002). From a food security perspective, any new technology that could enhance disease resistance without loss of crop yield should be explored fully. The work presented in this chapter aimed to identify inhibitors of AtHAM2, and homologues of AtHAM2 in crop species, in order to replicate the beneficial phenotypes of the *ham2* lines in a way that can be temporally controlled under field conditions.

4.6.1 MYST-family HATs are highly conserved

This chapter began with the identification of HAM2 homologues in other plant species. Using BLAST analysis, it was found that HAM2 retains high cross-species conservation, particularly within the economically important Brassicaceae family. Using *in silico* analyses, Cigliano et al. (2013) previously assessed the phylogeny of predicted HAM proteins from monocots (*Zea mays* and *Oryza sativa*) and dicots (*Solanum lycopersicum* and *Arabidopsis thaliana*). This analysis suggested that it was likely that a single ancestral *HAM* gene gave rise to HAMs in dicot and monocot species. These findings are encouraging, since compounds found to inhibit HAM2 activity in *Arabidopsis* and other dicot species may also inhibit HAM2 activity in economically important monocot species such as rice and maize. To assess HAM2 function, *Brassica oleracea* lines with HAM2 mutations mediated by CRISPR/Cas9 editing are currently being generated by collaborators. Once stable lines have been established, the morphological and disease resistance phenotypes of these plants will be assessed. Also, after inhibitors of HAM2 has been identified in *Arabidopsis*, they will also be tested on wild-type and *ham2* mutant *Brassica oleracea* lines.

4.6.2 Molecular modelling of AtHAM2, BnHAM2, SlHAM2 and AtHAG1

Homology models of the MYST-family members AtHAM2, BnHAM2, SlHAM2 and the GNAT-family member AtHAG1 were generated in a non-bound state. All HAM2

models were based on the solved structure of *Homo sapiens* KAT8 (PDB ID: 4DNC; resolution = 2.05 Å) and AtHAG1 was based on the solved structure of *Saccharomyces cerevisiae* GCN5 (PDB ID: 1YGH; resolution = 1.9 Å). All models were optimised by employing the maximum level of energy minimisation as permitted by MODELLER protocols, with these procedures being repeated twice for each model (see Section 2.7.2). By ranking according to Discrete Optimized Protein Energy (DOPE) (Shen and Sali, 2006) and Molecular Potential Density Function (molpdf) (Sali and Blundell, 1993) scores, the models with the most stable conformation, i.e. those with net inter-atomic force on each atom as close to zero as possible, were identified. Further analysis of backbone dihedral angles (Φ and Ψ) showed models were in good agreement with experimentally determined values (Lovell et al., 2003). Also, comparing residue-by-residue energies of each candidate model allowed the identification of erroneous or problematic high-energy regions, indicating unsuitability of that model. These efforts were taken to ensure that the models used for *in silico* docking were as close to the native protein structure as possible.

Homology models can also be structurally validated by *in silico* docking a library of known inhibitors and "decoy" random compounds. If the model is accurately close to the actual structure, the known inhibitors should bind with a distinctively higher docking score than the "decoy" compounds. The present study is the first to search for a MYST-family and plant-selective inhibitor; therefore there are no plant MYST inhibitors published in the literature and it was not possible to use this technique in this study. As a result of this work, and work following this, the identification of plant HAM-specific inhibitors will allow these analyses to be made in the future. In order to verify the models, each protein structure would be empirically solved (for example by X-ray diffraction or NMR), then the models and solved structures could be compared to determine the model accuracy.

4.6.3 Identification of inhibitors

Several steps were taken before commencing the *in silico* docking screen in order to improve the likelihood of identifying a successful inhibitor *in planta*. Firstly, to ensure that candidate compounds were likely to be foliar-absorbed and cell-permeable, a literature search of the properties of successful agrochemicals was performed. Using this information, compounds that were unlikely to perform as effective agrochemicals were removed from the screening library. Also, the performance and

suitability of AutoDock Vina was validated by successfully re-docking Ac-CoA back into a solved structure of HsKAT8 co-crystallised with Ac-CoA (Fig. 4.4.1).

The candidate compounds presented in Table 4.4 were initially identified for having higher docking scores against AtHAM2 than HsKAT8. The compounds also showed similar docking scores and poses against BnHAM2 and SlHAM2, and did not dock within the binding pocket of AtHAG1. As presented in the 2D interaction diagrams depicting the candidates docked into AtHAM2 (Fig. 4.12), the compounds show functional and stereochemical complementarity to the binding site. We were able to identify compounds with these desired properties through use of the novel application Sifter. By directly comparing the interactions made by Ac-CoA and AtHAM2 with the interactions made by the candidate compound and AtHAM2, Sifter was able to identify compounds that contain functional groups which mimic those of Ac-CoA (as described in Section 4.4.7). By adopting this technique, the candidate compounds identified as a result of this *in silico* screen should act as competitive inhibitors for the natural ligand binding site of AtHAM2, BnHAM2 and SlHAM2 *in planta*.

4.6.4 Limitations and further work

AutoDock Vina, the docking algorithm utilised in this study, allows test compound rotatable bonds to be flexible but keeps the protein structure static and rigid. It is possible to allow selected protein side-chains to be flexible, but these options increase the time and computational power required to perform calculations hugely. With a screening library of almost 1 million compounds, static docking was the only feasible option with the available resources and time during this study. It should also be noted that static docking algorithms lack explicit water treatments which are fundamental for reproducing specific protein-ligand complexes. Another limitation of this approach is that it provides only a static representation of the binding process, meaning that (1) the intrinsic dynamic nature of the protein is lost and any adaptation upon binding is not captured and (2) kinetic quantities cannot be estimated (Gioia et al., 2017). After short-listing candidate compounds based on static docking findings, it would be ideal to incorporate molecular dynamics (MD) techniques. With these methods, simulations of the physical movements of atoms are studied giving approximations of the dynamics of a system. With the inclusion of water molecules, MD would allow a more thorough exploration of AtHAM2-

inhibitor recognition and binding from a energetic and mechanistic point of view.

As well as MD simulations, the future directions of this work should be based on empirical testing of candidate inhibitors. Firstly, the 7 candidates shown in Table 4.4 will be applied to *Arabidopsis* Col-0, *ham2-71* and *ham2-75* seedling roots (along with NU9056 as a positive control) and the size of the root apical meristem zones will be assayed. Also, as mentioned previously, candidate inhibitors that show promising results in *Arabidopsis* will be progressed for testing on wild-type and *ham2* mutant *Brassica oleracea* lines. *In vitro* HAT assays are also appealing: if all of the twelve *Arabidopsis* HATs were cloned and purified, MYST-specific inhibitors could be identified using these assays. With the purified proteins, SPR (surface plasmon resonance) experiments could be performed allowing the determination of equilibrium dissociation constants from association and dissociation rates.

Altogether, the *in silico* work presented in this chapter will be followed with the *in planta* verification of MYST-specific inhibitors, leading to the development of a novel agrochemical based on the beneficial phenotypes observed in *Arabidopsis ham2* mutants.

Chapter 5

Investigating the Role of HATs in Effector-Triggered Immunity

5.1 Introduction

As described in Section 1.3.6, $\approx 10\%$ of the *Arabidopsis* genome becomes differentially expressed after bacterial perception (Zipfel et al., 2004; Denoux et al., 2008; Lewis et al., 2015). An outstanding yet fundamental question is how the host cell translates perception of an elicitor into an appropriate response through differential expression of this subset of the genome. One possible mechanism is through localised chromatin remodelling to modulate gene transcription (Wegel and Shaw, 2005). The accessibility of specific DNA regions to regulatory proteins, such as transcription factors and RNA polymerase machinery, is tightly controlled by chromatin-associated proteins such as histone-modifying enzymes and ATP-dependent chromatin remodelling complexes (Wegel and Shaw, 2005; Jerzmanowski and Archacki, 2017). Histone acetyltransferases (HATs), known to be key regulators of gene transcription (Kouzarides, 2007), transfer acetyl groups onto specific lysine residues of the N-terminus of histone tails which protrude outward from the nucleosome. This modification can affect chromatin structure directly by changing DNA/histone interactions within the nucleosome or indirectly by recruiting histone-modifying complexes (Kouzarides, 2007; Hsieh and Fischer, 2005). Given the key role of HATs in transcriptional regulation, an initial hypothesis was proposed in which HATs function in the regulation of transcriptional reprogramming that occurs when immune responses are elicited.

5.1.1 HAG1 is a positive regulator of basal defence

To investigate this hypothesis, a reverse genetics approach was previously taken in the group to assess the basal defences of T-DNA insertion histone acetyltransferase (*hat*) mutants to the hemibiotrophic bacterial pathogen *Pseudomonas syringae* pv. *tomato* (*Pto*) DC3000. Homozygous lines for 10 of the 12 *Arabidopsis* HATs were tested for altered immunity by infecting the lines with *Pto* DC3000 and quantifying bacterial growth 3 days post inoculation. The flagellin insensitive mutant *fls2* is known to be more susceptible than Col-0 to *Pto* DC3000 and was included as a control (Zipfel et al., 2004). Of the 10 *hat* mutants tested, one line (*hag1-6*) displayed increased susceptibility whilst another line (*ham2-75*) showed enhanced resistance to *Pto* DC3000 (Fig. 3.1). Whilst the previous two chapters focused on the more resistant *ham2* lines, the present chapter explores the role of HAG1, as well as the other *Arabidopsis* HATs, in immunity by investigating *hat* mutant defence responses.

As presented in Figure 5.1A, the T-DNA insertion in *hag1-6* lines is positioned within the first exon of *HAG1*. The insertion in the *hag1-6* line is predicted to cause a truncation such that the bromodomain and acetyltransferase domains are rendered non-functional. The *hag1-6* mutant line has been used and published in several studies including Long et al. (2006); Kornet and Scheres (2009); Servet et al. (2010); Chen et al. (2017). The acetyltransferase domain catalyses the transfer of an acetyl group to lysine residues, whereas the bromodomain has been implicated in targeting acetyltransferase activity through selective recognition and binding of acetylated lysine residues (Fujisawa and Filippakopoulos, 2017). The morphological phenotype of *hag1-6*, as observed in the literature by Servet et al. (2010), is presented in Figure 5.1B. The *hag1-6* mutant has a severe morphological phenotype, with much smaller leaf size than Col-0. The *hag1-6* mutant also has shorter roots, delayed flowering and is unable to produce viable seeds (Servet et al., 2010).

Previous work in the group also established that levels of histone acetylation at H3K9 (which HAG1 is known to target for acetylation *in vitro* (Earley et al., 2007)) increased within 5-10 minutes in Col-0 seedlings after elicitation with the bacterial PAMP flg22. The flagellin peptide flg22 is used to exclusively elicit PTI defence responses (no effectors are delivered to interfere with host responses). In *hag1-6* lines this effect was significantly reduced with very little induction of H3K9 acetylation at all time-points following flg22 elicitation (Piquerez et al., in preparation). Interestingly, the reduction of H3 acetylation correlated well with the susceptibility

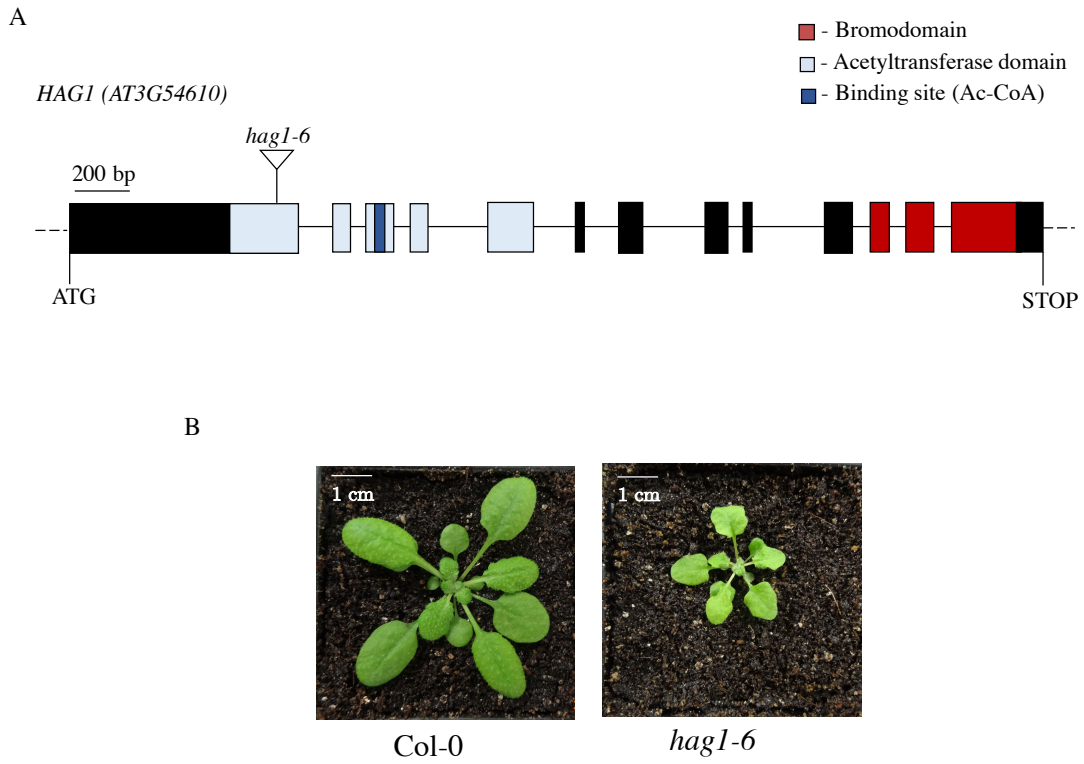


Figure 5.1: T-DNA insertion position and morphological phenotype of the *hag1-6* line. (A) T-DNA insertion position along the *HAG1* gene. Bars indicate position of exons, lines indicate introns and untranslated regions. (B) Observed morphological phenotypes of 4 week old *hag1-6* plants.

phenotype of *hag1-6*. To identify the specific genes that are acetylated in Col-0 after flg22, and to determine the effect of the *hag1-6* mutation on acetylation levels at these loci, ChIP-seq experiments are currently being conducted within the research group.

Further experiments also discovered that reactive oxygen species (ROS) production and mitogen-activated protein kinase (MAPK) activation was not significantly affected in *hag1-6* lines compared to Col-0. To test whether the susceptibility phenotype of *hag1-6* could be explained by impaired salicylic acid (SA) responses, *hag1-6* was crossed with *sid2-1* (SA-Induction Deficient 2 (*SID2*), a SA-biosynthetic gene) and *NahG* (a bacterial encoded gene for a SA-degrading salicylate hydroxylase) lines to lower SA levels (Delaney et al., 1994; Wildermuth et al., 2001). The double mutants displayed an enhanced susceptibility to *Pto* DC3000 phenotype, suggesting that the role of HAG1 in plant immunity does not overlap with SA-related

processes which are affected in the two SA mutant lines (Piquerez et al., in preparation). Overall, these findings support a model in which histone acetylation by HAG1 is a rapidly-induced output of PAMP-triggered immunity (PTI) activation which is downstream of ROS production and MAPK signalling, and is not related to SA-processes.

5.1.2 Effector-triggered immunity

A previously unexplored hypothesis is that HAG1, and other HATs, are also involved in effector-triggered immunity (ETI). As discussed in Section 1.3.3, plants have several levels of defence which are mostly effective against colonisation attempts by phytopathogens. The first inducible layer of defence is through extracellular perception of conserved pathogen-associated molecular patterns (PAMPs) by membrane-associated pattern recognition receptors (PRRs) (Dodds and Rathjen, 2010). PAMP perception by PRRs and subsequent defence signalling, known as PTI, consists of an array of molecular responses to limit pathogen growth. Virulent pathogens, including *Pto* DC3000, have adapted to suppress PTI by secreting effector proteins into host cells via the needle-like type 3 secretion system (T3SS) (Büttner and He, 2009). Effector proteins function to manipulate the structure and function of host cell components in order to suppress immunity for successful infection (Alfano and Collmer, 2004).

Plants have evolved the ability to detect pathogenic effectors in a gene-for-gene manner, where resistance (R) proteins (encoded by host disease resistance (*R*) loci) directly or indirectly recognise pathogenic effectors (encoded by avirulence (*avr*) loci in the pathogen) (Flor, 1971). The majority of *R* genes encode polymorphic NB-LRR (nucleotide-binding and leucine-rich repeat) proteins, which are subdivided into two classes based on the N-terminal domain. One class carries a TIR-domain, the other class carries a coiled-coil (CC) domain: these two distinct domains define specific signalling pathways that NB-LRRs initiate in response to effector recognition (Aarts et al., 1998).

The cellular events following R-protein perception of an effector are collectively known as effector-triggered immunity (ETI). Reactions include ROS burst, calcium ion influx, SA synthesis, activation of MAPK cascades, transcriptional up-regulation of defence genes and the hypersensitive response (HR) (Dodds and Rathjen, 2010).

HR is the rapid and localised programmed cell death (PCD) in infected cells in order to limit further biotrophic pathogen invasion. There are many similarities between the PTI and ETI responses. For example, the transcriptional changes induced by both defence strategies have been found to overlap extensively (Navarro et al., 2004). In fact, both responses rely on activation of MAPK cascades for signalling; silencing of *Arabidopsis* MPK6 was found to compromise resistance against both the virulent *Pto* DC3000 and the avirulent *Pto* DC3000 expressing *avrRpt2* (Menke et al., 2004). Overall, it is thought that the differences between PTI and ETI do not result from alternate downstream signalling machinery; instead the two modes of defence differ in duration and amplitude of response (Tsuda and Katagiri, 2010).

The involvement of several chromatin-associated regulatory proteins during the rapid transcriptional changes in ETI have been described in the literature. The *Arabidopsis* Elongator complex subunit 2 (ELP2) has been confirmed to be a positive regulator of ETI by affecting the levels of DNA methylation and induction of *NPR1* and *PAD4* genes in response to *Pto* DC3000 (*avrRpt2*) infection (Wang et al., 2013). Also a histone demethylase, *Arabidopsis* ATX1, was found to upregulate the expression of defence genes, such as *WRKY70*, the SA-responsive gene *PR1* and the JA-responsive gene *THI2.1*, through establishing the trimethylation pattern of H3 lysine 4 (H3K4me3) of its nucleosomes (Alvarez-Venegas et al., 2007).

As work on this study commenced, there were no publications in the literature describing a link between ETI responses and histone acetyltransferases. Recently, Kong et al. (2017) investigated how PsAvh23, an effector deployed by the soybean pathogen *Phytophthora sojae*, manipulates immunity in the host plant. It was found that PsAvh23 binds to the ADA2 subunit of the SAGA HAT complex, interfering with the necessary association with the catalytic subunit GCN5 (also known as HAG1 in *Arabidopsis*). This structural disruption caused a reduction of H3K9 acetylation levels, which are regulated by the ADA2/GCN5 module, at defence-gene loci and increased host colonisation by *Phytophthora sojae*. This study confirms the importance of histone acetylation and HAG1 in mounting immune responses, whilst also demonstrating that activity of the GCN5-containing SAGA complex represents an attractive target for phytopathogens.

5.1.3 Aims of this chapter

In this chapter, the role of *Arabidopsis* HATs (including HAG1) in ETI was explored through analysis of *hat* mutant ETI responses. The following strains were used in these investigations: *Pto* DC3000 empty vector (EV) (a virulent control), *Pto* DC3000 (*avrRpt2*), *Pto* DC3000 (*avrRpm1*), *Pto* DC3000 (*avrPphB*) and *Pto* DC3000 (*avrRps4*). AvrRpt2 and AvrRpm1 are well-studied *P. syringae* effectors that are recognised by the CC-NB-LRRs Resistant to *P. syringae* 2 (RPS2) and Resistant to *P. syringae* pv. *maculicola* 1 (RPM1) respectively. RPM1-interacting protein 4 (RIN4) plays a crucial role in perceiving both AvrRpt2 and AvrRpm1: RPM1 detects the phosphorylation of RIN4 by AvrRpm1 and RPS2 detects the cleavage of RIN4 by AvrRpt2, triggering ETI (Mackey et al., 2002, 2003). AvrPphB, a cysteine protease, cleaves AvrPphB susceptible 1 (PBS1) protein which is guarded by the CC-NB-LRR Resistant to *P. syringae* 5 (RPS5) (Shao et al., 2003) and AvrRps4 is detected by the paired TIR-NB-LRRs Resistant to *P. syringae* 4 (RPS4) and Resistant to *Ralstonia solanacearum* 1 (RRS1) to initiate downstream ETI signalling and cellular responses (Gassmann et al., 1999). By using several different effectors that are recognised by both CC- and TIR-NB-LRRs, we aimed to analyse the role of HATs in NB-LRR-specific signalling pathways following effector perception.

5.2 Assessing AvrRpt2-induced responses in *hat* mutants

5.2.1 AvrRpt2-induced ion leakage is attenuated in *hac4*, *hac5*, *ham1* and *hag1* mutants

To test whether HATs function in ETI responses after host recognition of a pathogen effector, T-DNA insertion mutants of 9 of the 12 *Arabidopsis* HATs were subjected to ion leakage assay (method described in Section 2.3.4). For this assay, hypersensitive response (HR)-dependent cell death following effector recognition is quantitated by measuring ion leakage from infected leaves. The mutant lines *hag1-6*, *hag2-99*, *ham1-96*, *ham2-71*, *hac2-84*, *hac4-21*, *hac5-78*, *hac12-04* and *haf2-29* were infiltrated with *Pto* DC3000 (EV) and *Pto* DC3000 expressing the type III effector *avrRpt2* (from here these strains are called *Pto* (EV) and *Pto* (*avrRpt2*). *Pto* (EV), which harbours the pVSP61 "empty vector" plasmid, was used to ensure

that the plasmid does not interfere with *Pto* growth. Figure 5.2 presents the results of this experiment; the control strain *Pto* (EV) induced minimal ion leakage in all genotypes tested, indicating that no effectors secreted by this strain induce HR. Consistent with the literature (Mackey et al., 2003), inoculation with *Pto* (*avrRpt2*) induced ion leakage from 6 hours post infiltration in Col-0. The mutant lines *hag1-6*, *hac4-21*, *hac5-78* and *ham1-96* were found to have a reduced ion leakage phenotype compared to Col-0 at one or more time-points during the assay. *hag1-6* showed the most severe impairment of AvrRpt2-induced ion leakage from all of the *hat* mutants tested. Figure 5.2 presents the statistically significant results of the initial screen; the non-significant results can be viewed in Appendix D. In Figure 5.3, the T-DNA insertion positions for the *hat* lines with ion leakage phenotypes, as well as other lines used in the next section, and the morphological phenotypes are presented. It can be seen that *hac4-21*, *hac5-78* and *ham1-96* have a similar morphological phenotype to Col-0.

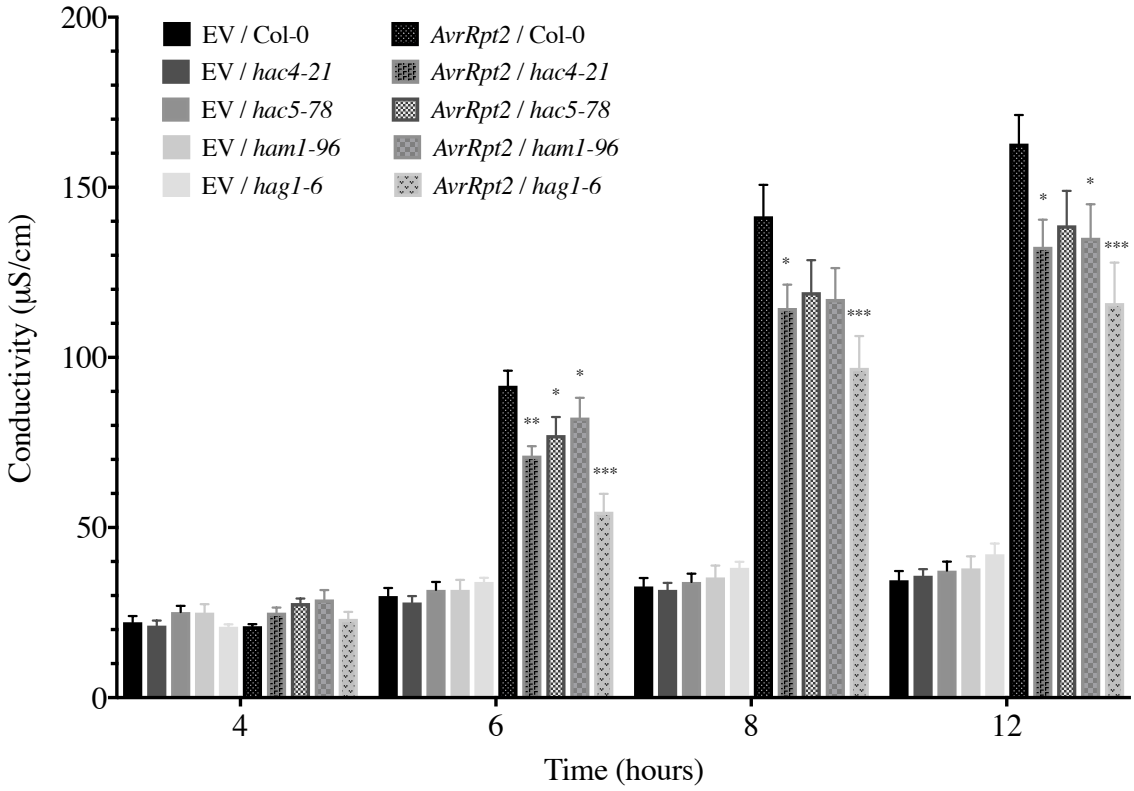


Figure 5.2: *hat* mutants *hac4*, *hac5*, *ham1* and *hag1* have compromised AvrRpt2-induced HR. Leaves of 5 week old plants were infiltrated with either *Pto* EV (empty vector) or *Pto (avrRpt2)* at OD_{600}= 0.1. Ion leakage was measured 4-12 hours after inoculation. Statistical significance versus Col-0 at each time-point determined by two-tailed t-test, n = 6, *** P $\leq$ 0.001, ** P $\leq$ 0.01, * P $\leq$ 0.05. Data shown are representative of 2 independent experiments. Error bars indicate standard error.

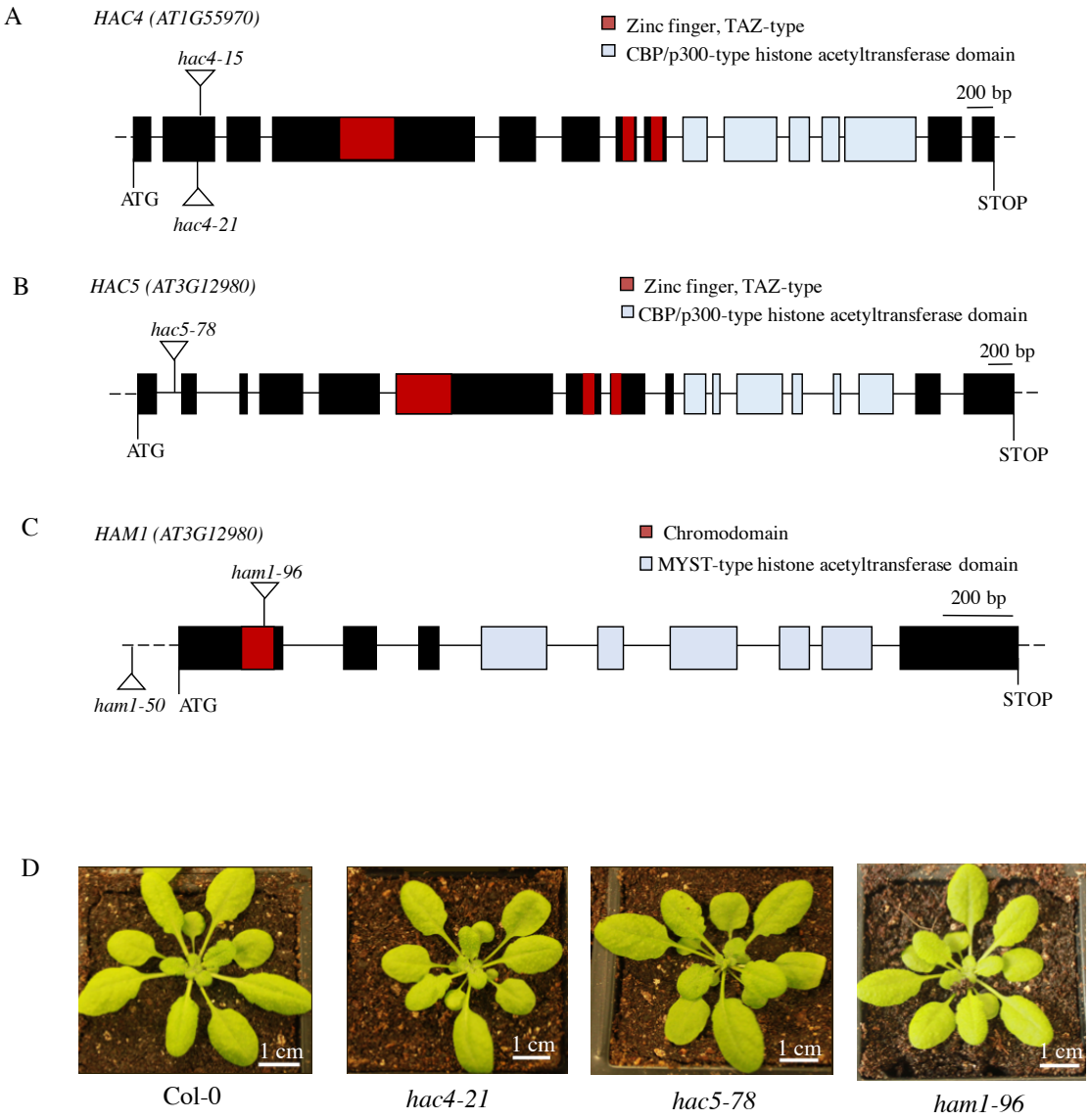


Figure 5.3: T-DNA insertion positions and morphological phenotypes of *hac4-21*, *hac5-78* and *ham1-96* lines. T-DNA insertion positions along the (A) *HAC4* (B) *HAC5* and (C) *HAM1* gene. Bars indicate position of exons, lines indicate introns and untranslated regions. (D) Observed morphological phenotypes of 4 week old *hac4-21*, *hac5-78* and *ham1-96* plants.

5.2.2 AvrRpt2-immunity is not affected in *hac4*, *hac5* and *ham1* mutants, but is in *hag1-6*

To determine whether the impairment in AvrRpt2-induced ion leakage leads to compromised AvrRpt2-triggered immunity in *hat* mutant lines, *in planta* bacterial growth assays were performed using *Pto* (EV) and *Pto* (*avrRpt2*). One extra T-DNA insertion line was available for *HAC4* and *HAM1* and were used in this experiment to confirm any phenotypes seen. Leaves of 5 week old plants were infiltrated with *Pto* (EV) and *Pto* (*avrRpt2*) at OD_{600}= 0.001 and bacterial growth was measured 3 days post inoculation (Section 2.3.2). As expected, we observed reduced growth of the avirulent strain *Pto* (*avrRpt2*) compared to virulent *Pto* (EV) in Col-0. The growth of both *Pto* (EV) and *Pto* (*avrRpt2*) showed no statistical difference in *hac4-15*, *hac4-21*, *hac5-78*, *ham1-50* and *ham1-96* lines compared to Col-0 (Fig. 5.4). These results indicate that, although early HR responses are negatively affected by mutations in *hac4*, *hac5* and *ham1* genes, other ETI responses attenuating the growth of the avirulent strain remain intact. As described in Section 5.5.1, double mutant *hac4/hac5* and *ham1/ham2* lines were created to uncover any potential redundancy with these enzymes.

A statistically significant increase in the growth of both *Pto* (EV) and *Pto* (*avrRpt2*) can be seen in *hag1-6* compared to Col-0 (Fig. 5.4). The avirulent strain *Pto* (*avrRpt2*) grew less than *Pto* (EV) on *hag1-6*, indicating that ETI processes are still active in *hag1-6* plants. To investigate ETI in *hag1-6* further, experiments were conducted using other *Pto* (*avr*) strains.

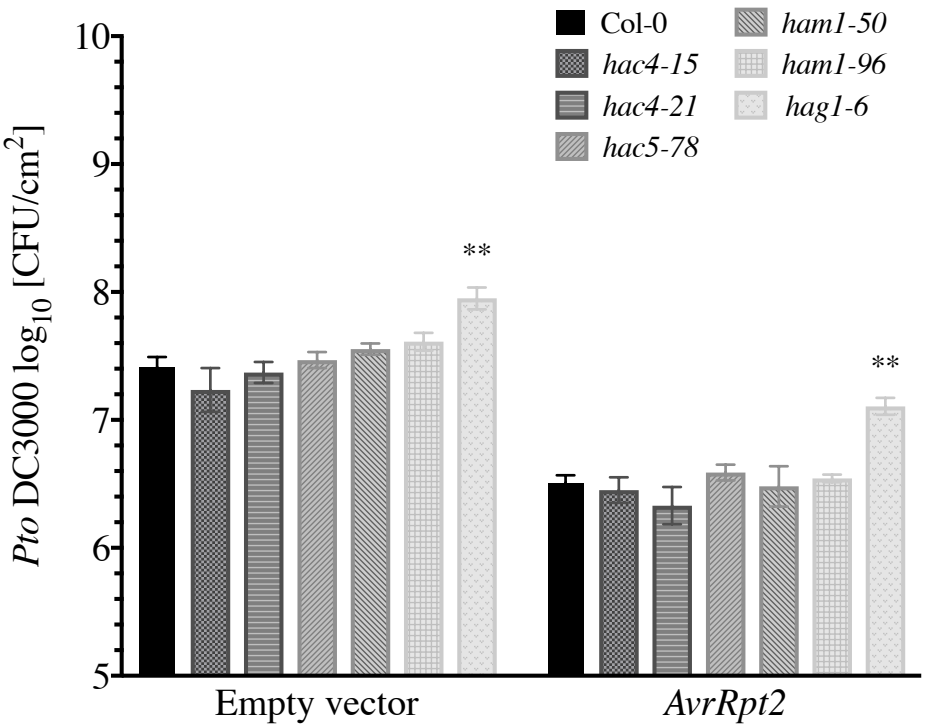


Figure 5.4: *Pto* (EV) and *Pto* (*avrRpt2)* growth is affected in *hag1*, but is not affected in *hac4*, *hac5* and *ham1* mutants. *Pto* (EV) and *Pto* (*avrRpt2*) (OD_{600}= 0.001) were infiltrated into leaves of 5 week old Col-0, *hac4-15*, *hac4-21*, *hac5-78*, *ham1-50*, *ham1-96* and *hag1-6*. Bacterial growth was measured 3 days post inoculation (Section 2.3.2). Data shown are representative of 3 independent experiments. Statistical significance versus Col-0 determined by two-tailed t-test, n = 6, ** $P \leq 0.01$. Error bars indicate standard error.

5.3 The role of HAG1 in ETI

5.3.1 AvrRpt2, AvrRpm1 and AvrPphB-induced ion leakage is attenuated in *hag1-6*

The results presented in this chapter suggest that HAG1 may function in establishing ETI after perception of AvrRpt2. To determine whether this effect was AvrRpt2-specific, ion leakage assays were performed on Col-0 and *hag1-6* lines after syringe infiltration with the avirulent strains *Pto* (*avrRpm1*) and *Pto* (*avrPphB*) (both are detected by CC-NB-LRR receptors). The remaining avirulent strain within our stocks that is detected by a TIR-NB-LRR receptor, *Pto* (*avrRps4*), does not induce ion leakage in Col-0 so therefore could not be used. Consistent with the literature, we observed that HR was more rapidly induced by AvrRpm1 (at 4 hpi) than both AvrRpt2 (at 6 hpi) and AvrPphB (at 8 hpi) in Col-0 (Ritter and Dangl, 1996). It can also be observed that ion leakage was impaired for all avirulent strains in *hag1-6*, and the most severe phenotype seen was for *Pto* (*avrPphB*). At 8 hours post infiltration onwards, ion leakage in *hag1-6* after inoculation with *Pto* (*avrPphB*) was less than half of that seen in Col-0. These results indicate that HAG1 may play a more dominant role in the distinct defence responses following AvrPphB perception than after AvrRpm1 and AvrRpt2 detection.

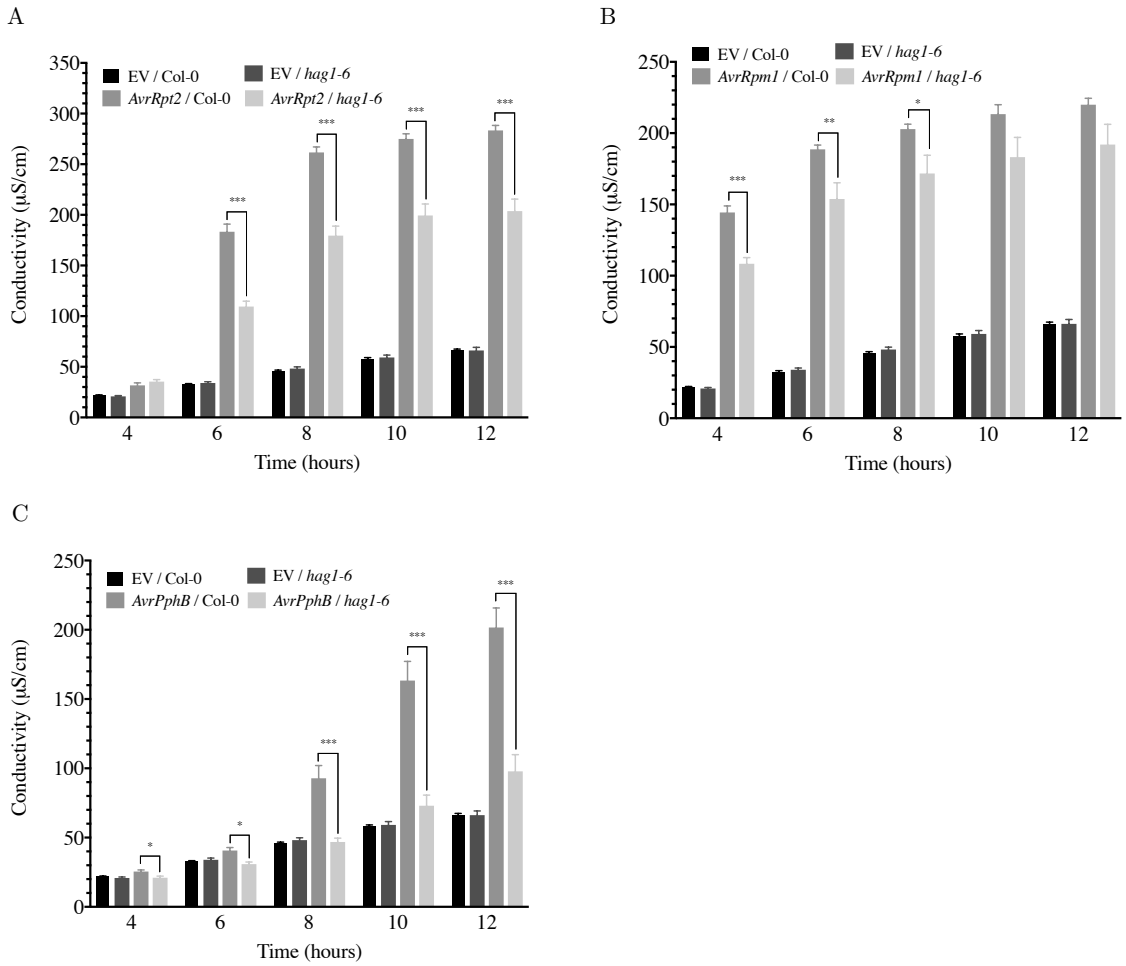


Figure 5.5: AvrRpt2-, AvrRpm1- and AvrPphB-induced HR is compromised in *hag1-6*. Leaves of 5 week old plants were infiltrated with either *Pto* (EV), *Pto* (*avrRpt2*), *Pto* (*avrRpm1*) or *Pto* (*avrPphB*) at OD_{600}= 0.1. Ion leakage was measured 4-12 hours after inoculation. Statistical significance versus Col-0 at each time-point and condition determined by two-tailed t-test, n = 6, *** $P \leq 0.001$, ** $P \leq 0.01$, * $P \leq 0.05$. Data shown are representative of 3 independent experiments. Error bars indicate standard error.

5.3.2 AvrRpt2, AvrPphB and AvrRps4-immunity is compromised in *hag1-6*

To determine if the attenuation of ion leakage seen for *hag1-6* in response to all effectors tested leads to compromised immunity against avirulent strains, *in planta* bacterial growth assays were performed using *Pto* (EV), *Pto* (*avrRpt2*), *Pto* (*avrRpm1*), *Pto* (*avrPphB*) and *Pto* (*avrRps4*). Figure 5.6 shows that *hag1-6* was

susceptible to all strains tested, virulent or avirulent, compared to Col-0. Growth of *Pto* (*avrRpm1*) was slightly enhanced in *hag1-6*, though it is not statistically significant compared with *Pto* (*avrRpm1*) growth in Col-0. It is interesting to note that *hag1-6* seems to be more susceptible to *Pto* (*avrRpt2*) and *Pto* (*avrPphB*), supporting the hypothesis that HAG1 may play a more prominent role in ETI responses following AvrPphB and AvrRpt2 detection compared to AvrRpm1 and AvrRps4. The question that still remains is if *hag1-6* is truly defective in ETI, or are the ion leakage and bacterial growth phenotypes presented in Figure 5.5 and Figure 5.6 due to the compromised basal defences of *hag1-6*.

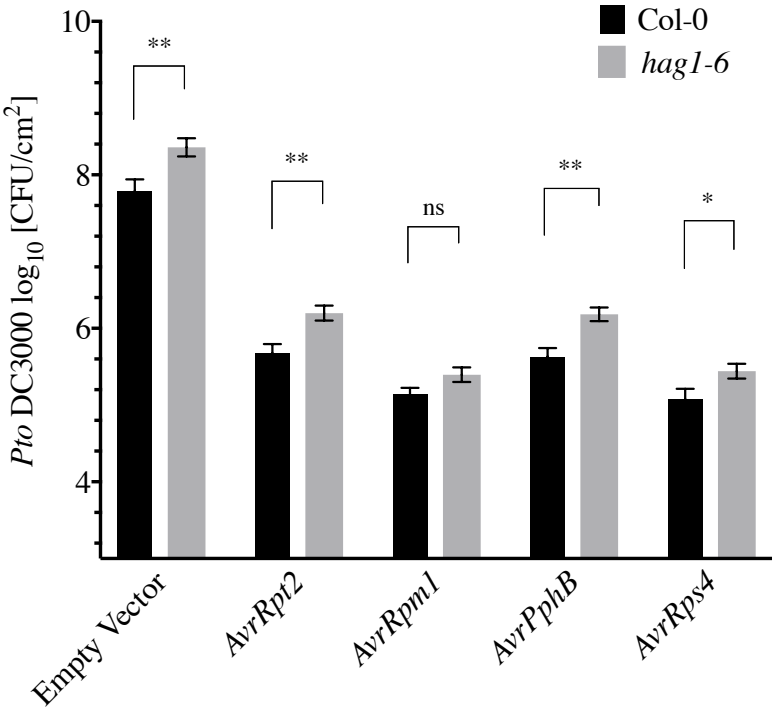


Figure 5.6: *hag1-6* is more susceptible to *Pto* (EV), *Pto* (*avrRpt2*), *Pto* (*avrPphB*) and *Pto* (*avrRps4*). Strains were infiltrated at OD_{600}= 0.001 into 5 week old leaves of Col-0 and *hag1-6*. Bacterial growth was measured 3 days post inoculation (Section 2.3.2). Statistical significance versus Col-0 for each condition determined by two-tailed t-test, n = 6, ** P $\leq$ 0.01, * P $\leq$ 0.05. Data shown are representative of 3 independent experiments. Error bars indicate standard error.

5.4 Transcriptomic analysis of *hag1-6* ETI responses

5.4.1 ETI marker gene expression in *hag1-6*

In order to determine if there are transcriptional changes underlying the defective ETI responses in *hag1-6*, we investigated the transcriptional differences between Col-0 and *hag1-6* when infected with *Pto* (*avrRpt2*) and *Pto* (*avrPphB*). The time-point of 6 hours post inoculation was chosen as the ion leakage assays indicated that HR was induced by this time in Col-0 after infiltration with *Pto* (*avrRpt2*) and *Pto* (*avrPphB*). These strains were chosen as they produced the most distinctive phenotypes in bacterial growth and ion leakage assays (Figs. 5.5, 5.6). Firstly the expression of marker genes known to be induced during ETI, *AvrRpt2-induced gene 1* (*AIG1*), *AvrRpt2-induced gene 2* (*AIG2*) and *WRKY46*, was analysed.

AIG1 and *AIG2* were initially identified by Reuber and Ausubel (1996) who demonstrated that the genes exhibit *RPS2* and *avrRpt2*-dependent induction at early time-points following infection with *P. syringae* pv *maculicola* strain ES4326 carrying *avrRpt2*. Since then, the two genes have been used as marker genes indicating the activation of ETI in publications such as Ritter and Dangl (1996), Cheng et al. (2013) and Hung et al. (2014). *WRKY46* belongs to the WRKY family of TFs that are key transcriptional regulators of biotic and abiotic responses in plants (Rushton et al., 2010). Gao et al. (2013) identified *WRKY46* as an early RPM1- and RPS2-dependent marker gene in convergent ETI signalling when performing transcriptomic analysis of protoplasts expressing *avrRpm1* and *avrRpt2*. *WRKY46* has been used as an ETI marker genes in publications including He et al. (2006), Cui et al. (2013) and Cheng et al. (2013).

As demonstrated in Figure 5.7, *AIG1*, *AIG2* and *WRKY46* were strongly up-regulated in Col-0 6 hours post inoculation with *Pto* (*avrPphB*) and even more so for *Pto* (*avrRpt2*). Also, a small level of up-regulation can be see for *AIG1* and *WRKY46* in response to *Pto* (EV), indicating non-Avr factors can also weakly induce expression. In *hag1-6*, it can be seen that up-regulation of all ETI marker genes is greatly diminished compared to Col-0 post inoculation with *Pto* (*avrPphB*) and *Pto* (*avrRpt2*). The differential expression of *AIG1*, *AIG2* and *WRKY46* between *Pto* (EV) and *Pto* (*avr*) strains in *hag1-6* indicates that there is a small degree of Avr-dependent up-regulation, though this is far less than that in Col-0.

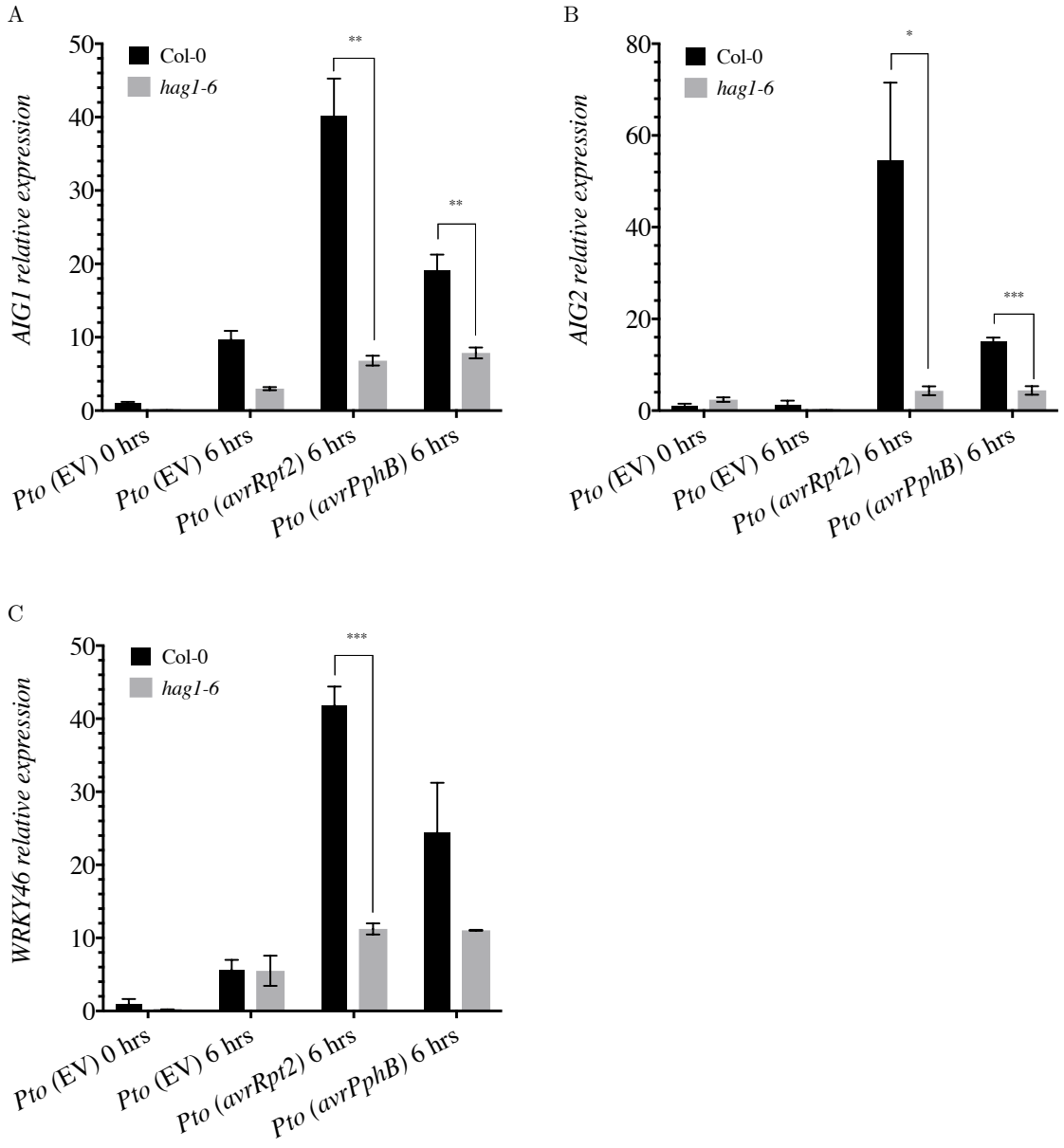


Figure 5.7: ETI marker gene expression is attenuated in *hag1-6* compared to Col-0. Relative expression of *AIG1*, *AIG2* and *WRKY46* in Col-0 and *hag1-6*. *Pto* (EV), *Pto* (*avrRpt2*) or *Pto* (*avrPphB*) (OD_{600}= 0.1) were infiltrated into 5 week old leaves and RNA was collected at designated time-points for qPCR. Relative expression values were determined using the comparative cycle threshold method ($2^{-\Delta\Delta Ct}$) and U-box housekeeping gene (AT5G15400) as a reference gene. Statistical significance versus Col-0 for each condition determined by two-tailed t-test, n = 3, *** $P \leq 0.001$, ** $P \leq 0.01$, * $P \leq 0.05$. Data shown are representative of 2 independent experiments. Error bars indicate standard error.

5.4.2 Microarray analysis of Col-0 and *hag1-6* in response to *Pto* (EV) and *Pto* (*avrRpt2*)

To assess the effect of the *hag1-6* mutation in response to virulent and avirulent infection at the transcriptome level in a more comprehensive way, microarrays were performed on 5 week old Col-0 and *hag1-6* plants post inoculation with *Pto* (EV) and *Pto* (*avrRpt2*). As presented in Figure 5.7, the ETI marker genes *AIG1*, *AIG2* and *WRKY46* were strongly induced 6 hours post infection with *Pto* (*avrRpt2*) in Col-0. This up-regulation was severely attenuated in *hag1-6* plants. Based on these previous findings, *Pto* (*avrRpt2*) was used as the avirulent strain for infection and samples were taken 6 hours post inoculation. Also, there are several publicly available microarray datasets of Col-0 infected with *Pto* (*avrRpt2*), allowing comparison of data (Zhang et al., 2013; Gu et al., 2016; Wang et al., 2014). Transcriptomic analysis was performed on the following samples:

Col-0

1. Mock (10 mM $MgCl_2$)
2. *Pto* (EV) (OD_{600}= 0.1)
3. *Pto* (*avrRpt2*) (OD_{600}= 0.1)

hag1-6

1. Mock (10 mM $MgCl_2$)
2. *Pto* (EV) (OD_{600}= 0.1)
3. *Pto* (*avrRpt2*) (OD_{600}= 0.1)

One RNA sample was extracted from 3 pooled leaves taken from different 5 week old plants (leaf number 7 was always chosen as this is the first fully formed adult leaf). Three biological replicates were performed on separate occasions at least one week apart. Experimental procedures were carried out by Sophie Piquerez (a post-doctoral research assistant in the group) and data analysis was performed by myself. A full description of experimental procedures and data processing can be found in Section 2.6.

Data normalisation

Quantile array normalisation was performed with the Bioconductor LIMMA (Linear Models for Microarray Data) software package in R (Ritchie et al., 2015). To ensure

the quality of the data, box-plots were produced to indicate the overall intensity of arrays; these are presented before normalisation and after normalisation in Figure 5.8. Here it can be seen that the data normalisation procedure was successful indicated by the alignment of all box-plots.

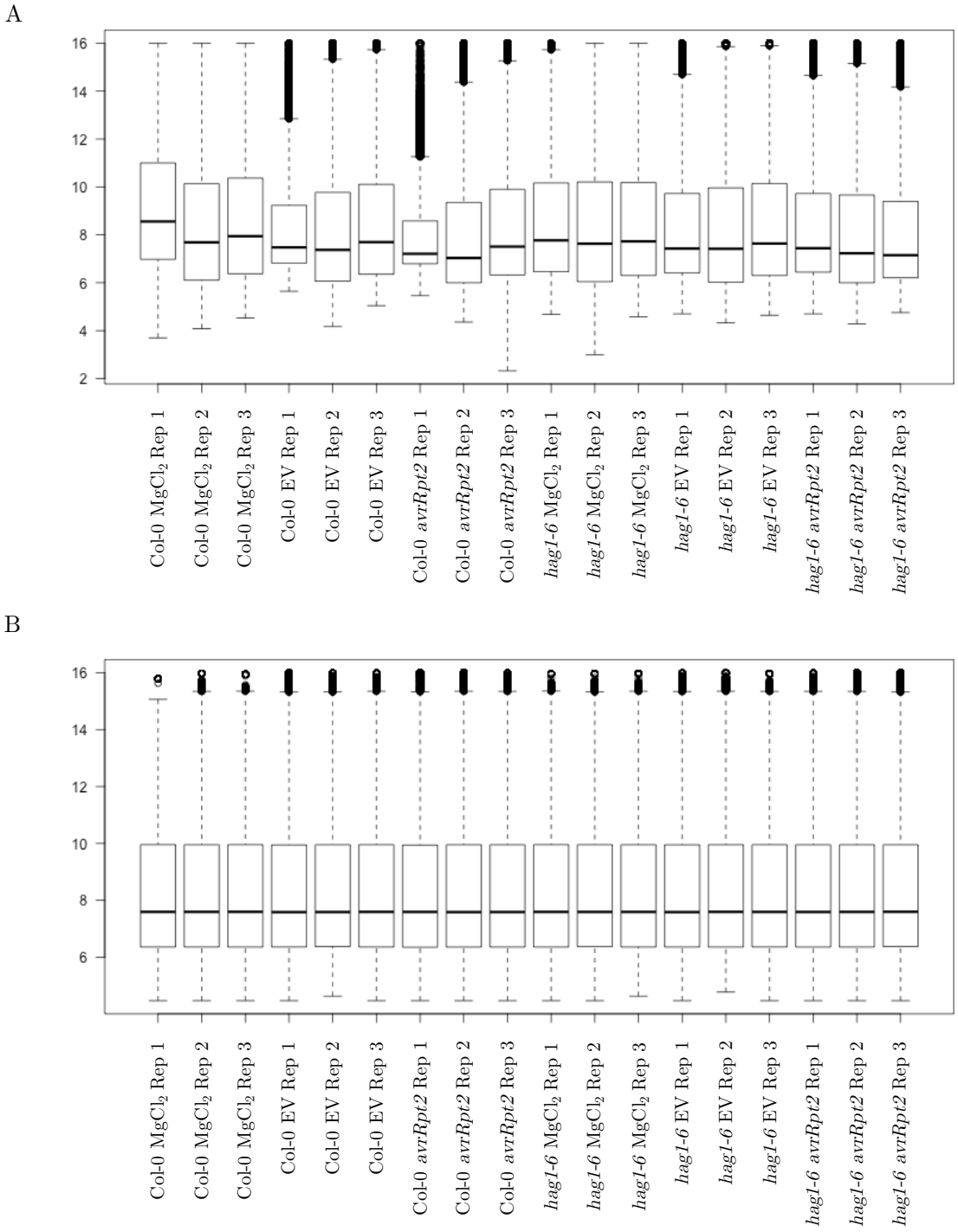


Figure 5.8: Overall microarray intensities before and after quantile normalisation. (A) Array intensities before normalisation. (B) Array intensities after normalisation. The centre of the box indicates the median, the box length indicates the interquartile range (IQR, 3rd quartile − 1st quartile), the upper whisker indicates the 3rd quartile + 1.5 * IQR, whereas the lower whisker indicates the 1st quartile 1.5 * IQR of the data.

Initial analysis of differentially expressed genes

Col-0

The LIMMA software package in R (Ritchie et al., 2015) was used to identify significantly differentially expressed genes (DEGs) between pairs of treatments using Benjamini-Hochberg false discovery rate (FDR) correction and thresholds of p-value ≤ 0.05 and $\log_2$ fold-change ≥ 1.5. Initial analysis showed that 1542 genes were up-regulated and 1965 genes were down-regulated in Col-0 in response to *Pto* (EV) infection (versus mock inoculation). These results were comparable to other published microarray data with *Pto* DC3000, such as Lewis et al. (2015), where 1562 genes up-regulated and 1694 down-regulated genes were identified.

When comparing gene expression after infection with *Pto* (*avrRpt2*) against mock, 3773 genes were up-regulated and 4247 down-regulated in Col-0. Results from this microarray experiment were found to be comparable with publicly available data from similar experiments (Table 5.1). The small differences between the publicly available expression data presented in Table 5.1 and the current data may be attributable to alternate sampling time-points, different *P. syringae* strains and different OD_{600} use for inoculations.

Table 5.1: The number of differentially expressed genes in publicly available microarray data utilising AvrRpt2 (Wang et al. (2014) (GEO accession: GSE58954) and Gu et al. (2016) (GEO accession: GSE72742)). Adjusted p-value ≤ 0.05 and $\log_2$ fold-change ≥ 1.5 thresholds were applied to identify differentially expressed genes between control and infected samples.

	Wang et al. (2014)	Gu et al. (2016)	Present study
Strain	*Psm* (*avrRpt2*)	*Psm* (*avrRpt2*)	*Pto* (*avrRpt2*)
OD	0.01	0.01	0.1
Time-points	0, 6 hpi	0, 6 hpi	6 hpi
Tissue	Leaf	Leaf	Leaf
Age	4 weeks old	4 weeks old	5 weeks old
Background	Col-0	Col-0	Col-0
Up-regulated genes (Mock vs *avrRpt2*)	3590	4698	3773
Down-regulated genes (Mock vs *avrRpt2*)	4311	5045	4247

hag1-6

When the same analysis was performed on *hag1-6* samples, results were notably different: 1342 genes were up-regulated and 1687 genes were down-regulated after infection with *Pto* (EV) versus mock. After infection with *Pto* (*avrRpt2*), 756 genes were up-regulated and 266 genes were down-regulated compared to mock in *hag1-6* (approximately 75 % less differential expression compared to Col-0 under the same conditions).

The differential expression between the 6 conditions is represented as a heat map in Figure 5.9. Here it can be seen that Col-0 and *hag1-6* mock inoculation (10 mM $MgCl_2$) data cluster together. It is interesting to note that, when infected with either *Pto* (EV) or *Pto* (*avrRpt2*), the data cluster based on genotype rather infection strain. This indicates that Col-0 and *hag1-6* plants respond very differently when treated with a virulent or avirulent strain, supporting the hypothesis that HAG1 plays an important role in host responses to Avr proteins.

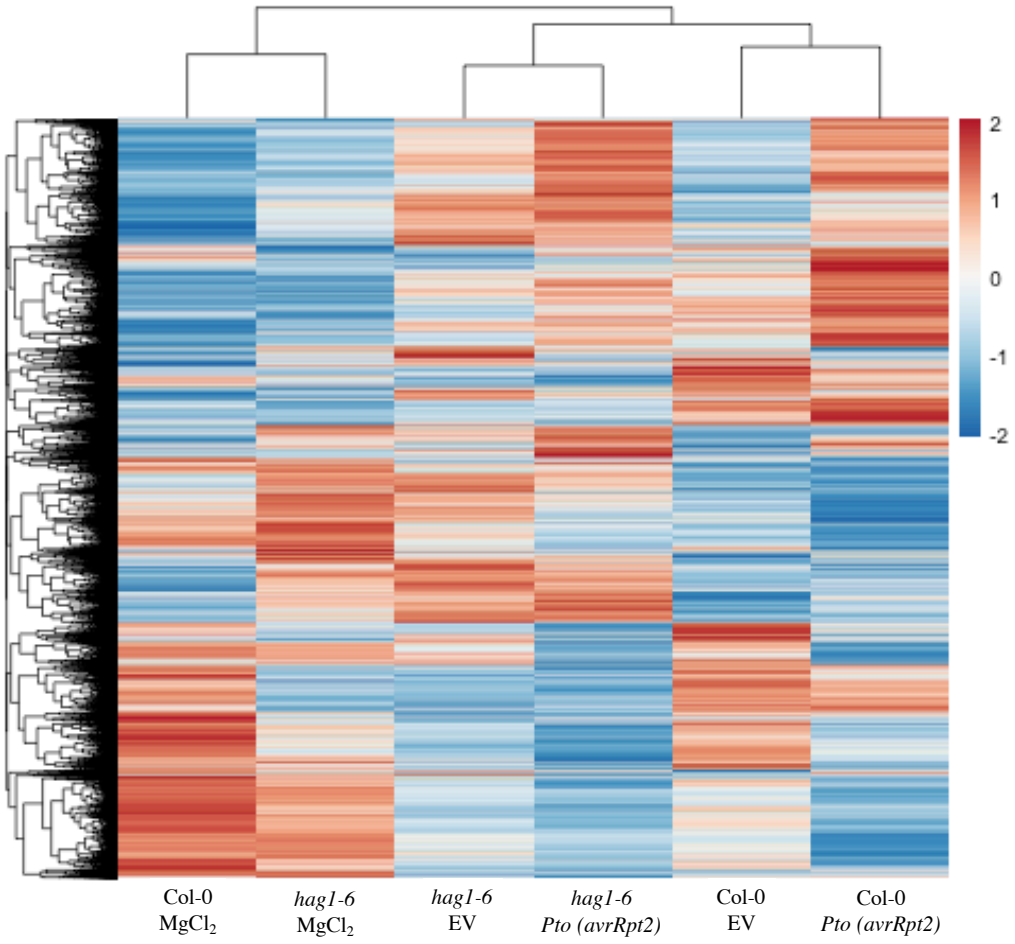


Figure 5.9: Heat map representation of differential gene expression between Col-0 and *hag1-6* after inoculation with 10 mM $MgCl_2$, *Pto* (EV) and *Pto* (*avrRpt2*). Relative expression values after array normalisation were subjected to hierarchical clustering. Red indicates high expression and blue indicates low expression.

Differentially expressed genes prior to infection

Differential expression analysis showed that, compared to Col-0, 803 genes had significantly greater basal expression levels and 1414 genes had significantly lower basal expression levels in *hag1-6* in mock samples (Fig. 5.10). BINGO analysis, as previously used in Chapter 3, is a tool which identifies statistically over-represented GO terms within a set of genes. This analysis was performed on the set of 803 genes with higher expression in *hag1-6* than Col-0 (with *Arabidopsis* Col-0 genome (TAIR10) as the reference population), and GO terms including protein transport (adjusted p=3.3e-5), cell cycle process (adjusted p=1.22e-3) and localisation in cell (adjusted p=1.22e-3) were over-represented. When this analysis was performed on the 1414

genes with lower expression in *hag1-6* than Col-0, GO terms related to metabolic processes, such as organic acid metabolism, amino acid/derivatives metabolism and small molecule biosynthesis were significantly over-represented. These findings suggest that metabolic and biosynthetic pathways could be down-regulated in *hag1-6*, whilst the enhanced expression of cell-cycle and chromosome segregation genes may induce increased cell cycle rates in this mutant.

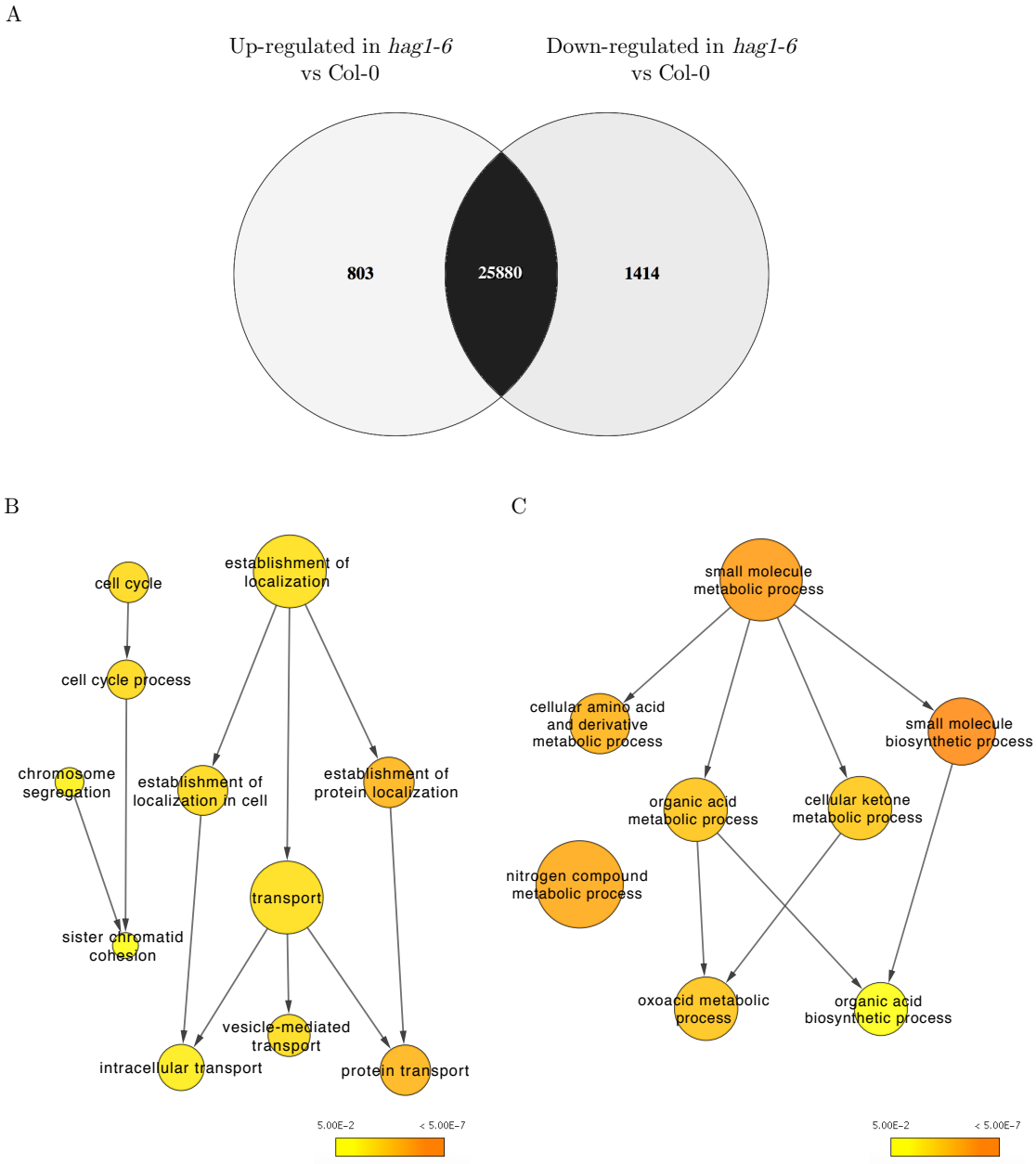


Figure 5.10: Differential expression genes between Col-0 and *hag1-6* under mock conditions. (A) Venn diagram representation of differential expression: 803 genes have greater expression and 1414 genes have reduced expression in *hag1-6* compared to Col-0. (B) BINGO visualisation of a selection of over-represented GO terms within the set of genes more expressed in *hag1-6* compared to Col-0. (C) BINGO visualisation of a selection over-represented GO terms within the set of genes less expressed in *hag1-6* compared to Col-0. Hypergeometric test with Benjamini and Hochberg FDR correction and threshold p-value ≤ 0.05 used to identify over-represented GO terms. Colour key indicates p-value of GO term.

Differentially expressed genes after *Pto* infection

To investigate the differences between Col-0 and *hag1-6* in their responses to *Pto* DC3000 (EV) and *Pto* (*avrRpt2*), scatterplots based on pairwise comparisons of gene expression between all conditions are presented in Figure 5.11. In these plots, the x-axis indicates the $\log_2$ fold change between *Pto* (EV) and mock, whereas the y-axis indicates the $\log_2$ fold change between *Pto* (*avrRpt2*) and mock. Each significantly differently expressed gene (based on thresholds of adjusted p-value ≤ 0.05 and $\log_2$ fold-change ≥ 1.5) is represented by a dot and is coloured based on its response to the three different treatments. Red represents genes that are either induced or repressed when infected with *Pto* (EV) compared to mock (virulent-responsive genes). Blue represents genes that are differentially expressed when infected with *Pto (avrRpt2)* compared to mock (avirulent-responsive genes). Genes that are differentially expressed in the same direction in response to *Pto* (EV) and *Pto (avrRpt2)* compared to mock are represented in green (viruent and avirulent-responsive genes). Purple represents genes that are differentially expressed in response to *Pto (avrRpt2)* compared to *Pto* (EV).

Differentially expressed genes after *Pto* (EV) infection

As expected (due to the basal susceptibility phenotype of *hag1-6*), differences between the number of genes responsive to *Pto* (EV) can be seen between Col-0 and *hag1-6* (red and green dots in Fig. 5.11). For Col-0, 3507 genes are induced or repressed in response to *Pto* (EV). In comparison, 3029 genes are differentially expressed in *hag1-6*. More specifically, it was found that 777 genes were up-regulated in Col-0 but not in *hag1-6*. To investigate the responses that *hag1-6* is defective in, BINGO analysis was performed on this set of 777 genes and over-represented categories included response to hormone stimulus (adjusted p=1.01e-5), regulation of transcription (adjusted p=6.69e-5) and regulation of macromolecule biosynthetic process (adjusted p=1.07e-4). These findings indicate that the susceptibility phenotype of *hag1-6* may be caused by dysregulation of hormonal responses, transcriptional processes and biosynthetic processes which are of critical importance for an effective defence response.

Differentially expressed genes after *Pto* (*avrRpt2*) infection

In the context of the current chapter where ETI responses are being explored, the most interesting results are the genes that are differentially expressed in response

to *Pto (avrRpt2)* compared to *Pto* (EV). In Col-0 it can be seen that, 6 hours post inoculation, 4096 genes are differentially expressed when comparing *Pto (avrRpt2)* and *Pto* (EV). When GO term analysis was performed on the genes up-regulated in response to *Pto (avrRpt2)* in Col-0, the most significantly over-represented categories included response to immune system process (adjusted p=1.59e-14), post-translational protein modification (adjusted p=2.00e-11) and programmed cell death (adjusted p=1.99e-6). This indicates that the transcriptional reprogramming to mount the effector-triggered hypersensitive response has occurred 6 hours post inoculation with *Pto (avrRpt2)* in Col-0. Analysis on the down-regulated set of genes found that categories including photosynthesis (adjusted p=7.97e-56), metabolic process (adjusted p=1.12e-25) and plastid organisation (adjusted p=1.4e-19) were significantly over-represented. These results are well correlated with publications which describe the targeting of chloroplasts by pathogen effectors to disrupt the chloroplastic reactive oxygen burst (de Torres Zabala et al., 2015; Lewis et al., 2015).

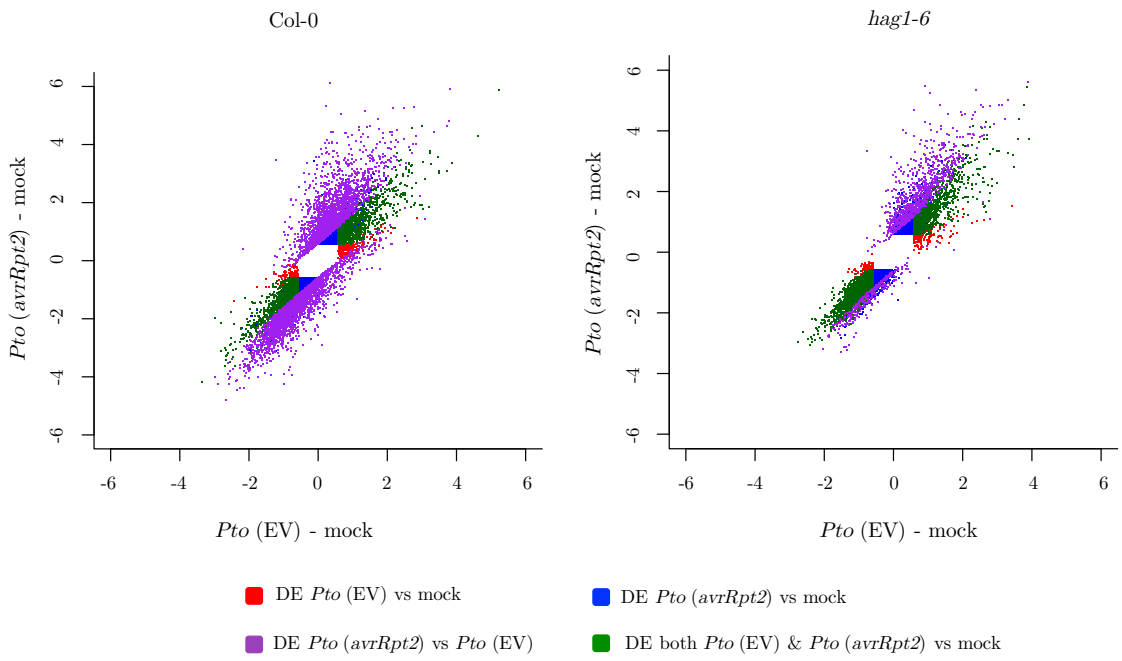


Figure 5.11: Differential expression between the Col-0 and *hag1-6* transcriptome when challenged with *Pto* (EV) and *Pto* (*avrRpt2*) represented by a scatterplot. The x-axis indicates $\log_2$ fold-change between *Pto* (EV) and mock, whereas the y-axis indicates $\log_2$ fold-change between *Pto* (*avrRpt2*) and mock. Red represents genes that are either induced or repressed when infected with *Pto* (EV) compared to mock. Blue represents genes that are differentially expressed when infected with *Pto (avrRpt2)* compared to mock. Green represents genes that are differentially expressed in the same direction in response to *Pto* (EV) and *Pto (avrRpt2)* compared to mock. Purple represents genes that are more differentially expressed in response to *Pto (avrRpt2)* than to *Pto* (EV). Differential expression analysis was performed using the Bioconductor LIMMA package using Benjamini-Hochberg false discovery rate (FDR) correction and thresholds of p-value ≤ 0.05 and $\log_2$ fold-change of ≥ 1.5.

Next, the analysis was extended to dissect the response of *hag1-6* to *Pto* (*avrRpt2*). Most of the genes differentially regulated in *hag1-6* in response to *Pto* (*avrRpt2*) were also differentially expressed in Col-0 under the same conditions. Interestingly, and importantly, more than 2000 genes were differentially expressed in Col-0 but were not in *hag1-6* in response to *Pto* (*avrRpt2*), as demonstrated by the Venn diagrams presented in Figure 5.12. This indicates that a large portion of the transcriptional responses of Col-0 to *Pto* (*avrRpt2*) does not occur in *hag1-6*.

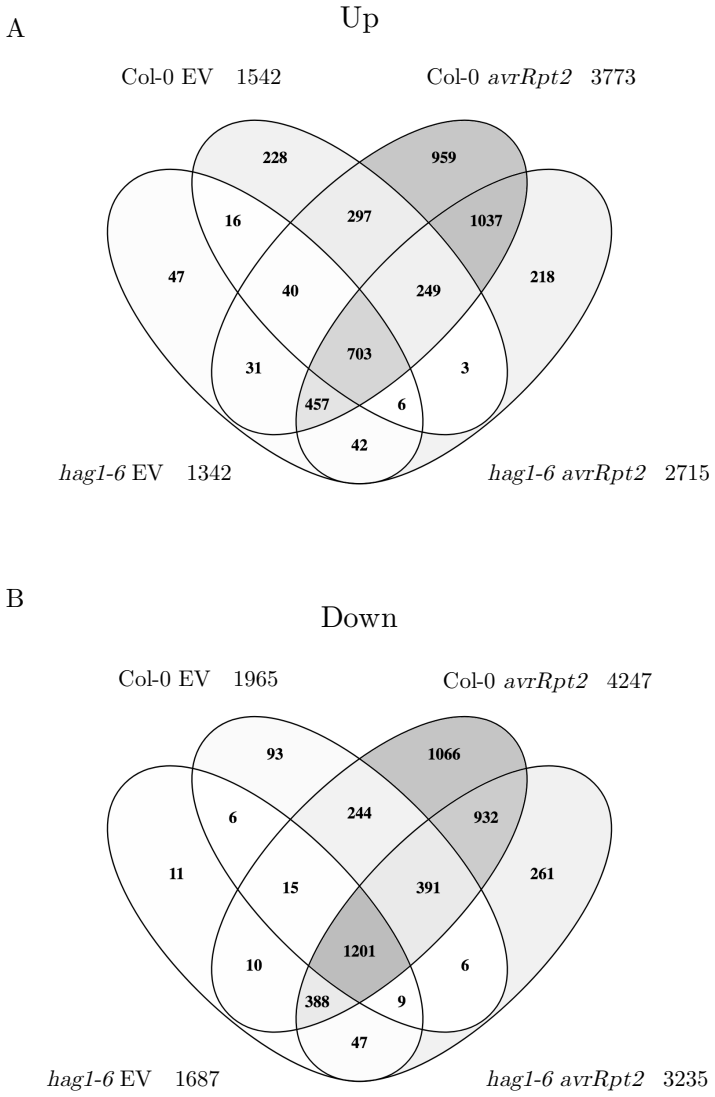


Figure 5.12: Venn diagrams of differentially expressed genes between Col-0 and *hag1-6* after infection with *Pto* (EV) or *Pto* (*avrRpt2*) compared to mock. The total number of differentially expressed genes for each condition is stated next to condition labels. (A) In Col-0, 1542 genes in EV and 3773 genes in *avrRpt2* were up-regulated compared to mock. In *hag1-6*, 1342 genes in EV and 2715 genes in *avrRpt2* were up-regulated compared to mock. (B) 1965 genes in EV and 4247 genes in *avrRpt2* were down-regulated in Col-0 compared to mock. 1687 genes in EV and 3235 genes in *avrRpt2* were down-regulated in *hag1-6* compared to mock.

In order to further characterise the transcriptomic deficiencies which *hag1-6* shows in response to *Pto* (*avrRpt2*), GO term analysis was performed on the 959 genes that

were up-regulated in Col-0 upon *Pto* (*avrRpt2*) infection but not in *hag1-6*. Also, the 1066 genes that are down-regulated in Col-0 upon *Pto* (*avrRpt2*) infection but not in *hag1-6* were subjected to GO term analysis. These results are presented in Figure 5.13. Over-represented categories in the 959 up-regulated genes included catabolic process (adjusted p=1.10e-5) and vesicle-mediated transport (adjusted p=1.47e-4). Over-represented categories in the 1066 down-regulated genes included plastid-related processes (adjusted p=2.74e-19), metabolic processes (adjusted p=3.21e-5) and cellular amino acid and derivative metabolic processes (adjusted p=2.66e-3). These findings suggest that, in *hag1-6* after infection with *Pto* (*avrRpt2*), catabolic processes (i.e. the break-down of molecules into smaller units) are not mounted, and metabolic processes are not arrested. Overall, these results confirm that *hag1-6* is unable to mount ETI responses at the transcriptional level.

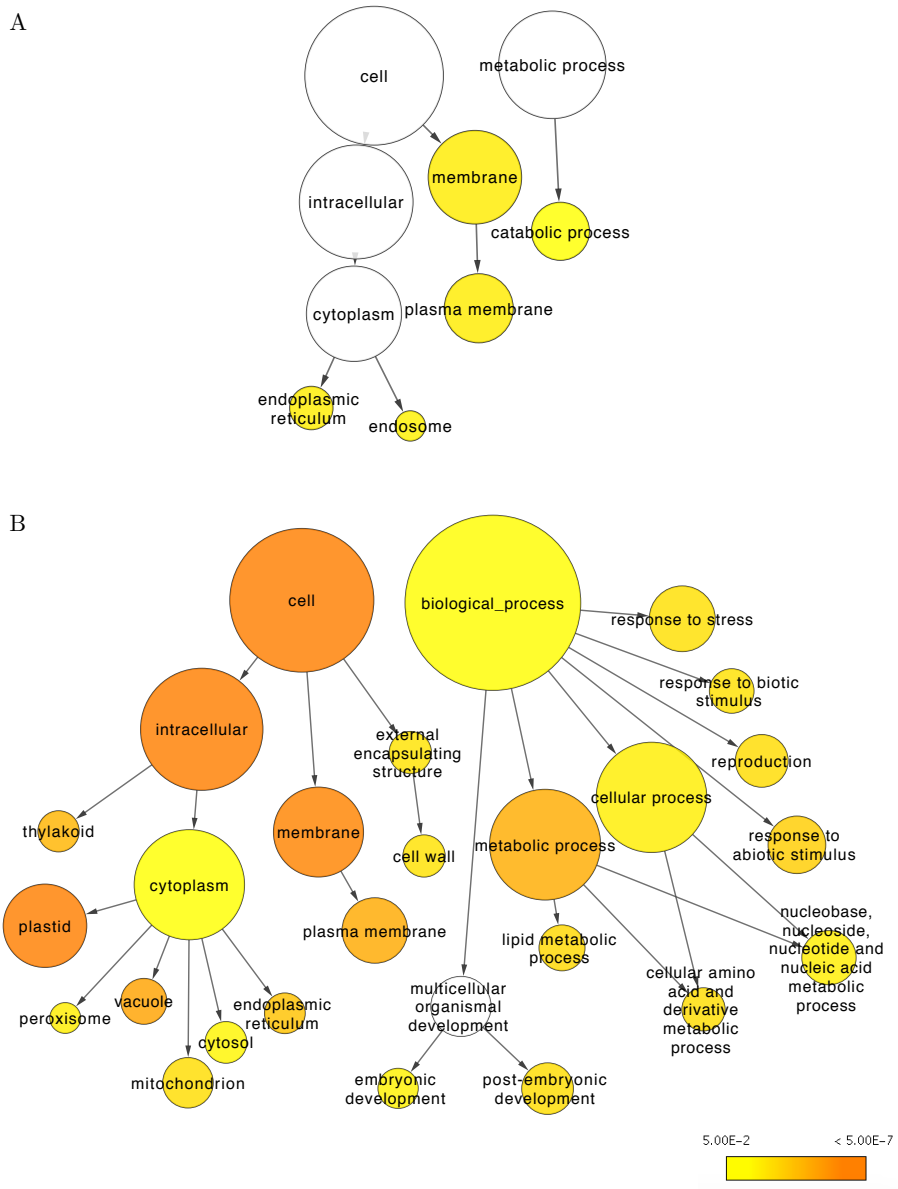


Figure 5.13: BINGO visualisation of the genes differentially expressed in Col-0 but not *hag1-6* upon infection with *Pto* (*avrRpt2*). (A) Over-represented GO terms within the set of 959 genes up-regulated in Col-0 upon *Pto* (*avrRpt2*) infection but not in *hag1-6*. (B) Over-represented GO terms within the set of 1066 genes down-regulated in Col-0 upon *Pto* (*avrRpt2*) infection but not in *hag1-6*. Hypergeometric test with Benjamini and Hochberg FDR correction and threshold p-value ≤ 0.05 used to identify over-represented GO terms. Colour key indicates p-value of GO term.

5.5 Discussion

5.5.1 AvrRpt2-induced responses of *hac4*, *hac5* and *ham1*

The initial screen for altered cell death responses uncovered a previously undiscovered role for *HAC4*, *HAC5* and *HAM1* in AvrRpt2-induced HR. We tested whether this attenuated cell death phenotype was linked to enhanced *Pto* (*avrRpt2*) growth in these *hat* mutants using *in planta* bacterial growth assays. The growth of both *Pto* (EV) and *Pto* (*avrRpt2*) showed no statistical difference in the *hat* mutants compared to Col-0 (Fig. 5.4), indicating other ETI responses attenuating the growth of the avirulent strain remain intact in these mutants.

As described in the main introduction (Section 1.5.1), *HAC4* and *HAC5* are members of the CBP family of *Arabidopsis* HATs which show homology to mammalian p300/CBP family of transcriptional co-activators (Pandey et al., 2002). In *Arabidopsis*, there are five p300/CBP HAT homologues with broad-specificity H3 lysine acetylation activity (Earley et al., 2007). Li et al. (2014a) demonstrated the within-family redundancy of *Arabidopsis HAC*s by analysing morphological and developmental phenotypes of single, double and triple *hac* mutants. They found that *HACs* are involved in pleiotropic developmental processes, with *HAC1* playing the most dominant role with the synergistic assistance from *HAC5*, *HAC4* and *HAC12*. We hypothesised that the functional redundancy of HAC4 and HAC5 caused the growth of *Pto* (*avrRpt2*) to be unaffected in single *hat* mutants compared to Col-0.

Although it has been published that the two *Arabidopsis* MYST HATs act redundantly (Earley et al., 2007; Latrasse et al., 2008), the ion leakage results presented here demonstrate non-redundant roles for HAM1 and HAM2, as *ham2-71* lines were unaffected in AvrRpt2-induced HR. The AvrRpt2-induced HR phenotype of *ham1-96* did not translate to an altered growth phenotype of *Pto* (*avrRpt2*) in *ham1-96*. Similar to *HAC4* and *HAC5*, we hypothesised that the functional redundancy of HAM1 and HAM2 may be masking AvrRpt2-immunity phenotypes in single *ham1* mutants compared to Col-0.

In light of the potential functional redundancy between HAC4 / HAC5 and HAM1 / HAM2, *hac4 / hac5* and *ham1 / ham2* double mutant lines were generated during this study. Further work with these lines will include ion leakage assays to assess if impaired cell death phenotypes are enhanced in these double mutants. Also

bacterial growth assays will be performed, using *Pto* harbouring other avirulence proteins such as AvrRpm1, to investigate differential responses of the mutant lines to alternate effectors.

5.5.2 ETI responses in *hag1-6*

The results presented in this chapter indicate that *hag1-6* is not just affected in basal defences as uncovered by previous work within the research group: the mutant is also defective in effector-triggered responses. The ion leakage results indicate that *hag1-6* does not respond as Col-0 does to AvrRpt2, AvrRpm1 and AvrPphB, suggesting that the HAG1 histone acetyltransferase functions at a point after effector-specific ETI signals are converged. The degree of response deficiency does vary between effectors in *hag1-6* (AvrRpt2-induced HR is more attenuated than AvrRpm1-induced HR (Figs. 5.5 and 5.6)). In the present chapter, we took a comprehensive look into AvrRpt2-induced transcriptional changes in Col-0 and *hag1-6*. In order to understand the phenotypic differences observed for *hag1-6* in response to different effectors, it would be very interesting to perform further transcriptomic characterisation of *hag1-6* when infected with *Pto* harbouring different ETI-inducing effectors.

The findings of the qPCR and microarray experiments presented in this chapter are in agreement: the expression of effector-induced genes 6 hours post infection with *Pto* (*avrRpt2*) was severely reduced in *hag1-6* compared to Col-0. Through the microarray analysis, we were able to dissect the different responses of *hag1-6* in the different phases of plant defence separately. In agreement with publications describing the developmental defects of *hag1-6* (Long et al., 2006; Kornet and Scheres, 2009; Servet et al., 2010; Chen et al., 2017), we confirmed that basal differences between Col-0 and *hag1-6* involved up-regulation of genes related to cell cycle processes, with metabolism-related genes being down-regulated.

The microarray analysis confirmed the ion leakage and bacterial growth assay results, which uncovered an unpublished role relating HAG1 activity to ETI defence responses. We found that approximately 2000 genes that were differentially expressed in Col-0 in response to *Pto* (*avrRpt2*) were not differentially expressed in the *hag1-6* mutant. This substantial transcriptional difference indicates that HAG1 is a key regulator of the genetic reprogramming that contributes to ETI cellular responses, such as the hypersensitive response. Mutating *HAG1* function therefore

leads to severe attenuation of the hypersensitive response, as presented in Figure 5.5. To completely isolate ETI responses, and remove the potentially interfering effects of other pathogen components, new plant lines expressing single effectors (including AvrRpt2 and AvrPphB) under a dexamethasone-inducible promoter were generated in a Col-0 and *hag1-6* background. With these lines, experiments such as qPCR analysis and ion leakage assays will allow us to simplify the ETI trigger in order to completely isolate the response of *hag1-6* to single pathogen effectors. It would also be interesting to assess the changes to H3K9 acetylation levels (the known target of HAG1) in these plant lines. This would allow us to determine if HAG1 alters the level of H3K9 acetylation upon AvrRpt2 perception at ETI-induced genes.

5.5.3 Limitations and further work

To investigate the mechanisms underlying HAG1 regulation of ETI, the following experiments will be performed in the future:

1. Using the transcriptomic data generated in this study, perform motif analysis on the set of genes not differentially expressed in *hag1-6* compared to Col-0 post infection with *Pto* (*avrRpt2*). In this way, potential transcription factor binding partners which guide HAG1 activity to specific genomic loci may be identified
2. Create *HAG1* complemented lines to restore the observed phenotypes seen for *hag1-6*
3. Create *HAG1*-tagged lines and perform ChIP to identify specific targets of HAG1 under different experimental conditions
4. The microarray and qPCR analysis performed here was just a snapshot of transcriptional activity: a time-course of earlier and later time-points would allow us to investigate the temporal differences between Col-0 and *hag1-6* ETI responses
5. Also, it would be ideal to analyse effector mRNA levels post infection to confirm equal amounts of the effector were delivered to each leaf. Ion leakage assays were performed with same leaf surface area, but cell patterning of *hat* mutants was not checked to account for morphological differences that would contribute to different infection phenotypes

Chapter 6

General Discussion

The breadth of literature relating HATs, HDACs and histone acetylation to a plethora of cellular processes demonstrates the importance of histone acetylation during the life cycle of a plant. Histone acetylation is fundamental to developmental, morphological, abiotic and biotic stress responses. The work presented in this thesis confirms that two histone acetyltransferases, HAG1 and HAM2, play important roles in defences against plant pathogens. Work presented in other publications and novel findings discovered in the present study have established that HAG1:

- Positively regulates growth and developmental processes (Vlachonasios et al., 2003; Benhamed et al., 2006; Kornet and Scheres, 2009)
- Positively regulates basal defence (Piquerez et al., in preparation)
- Positively regulates ETI - a novel finding resulting from this work (Chapter 5)

For HAM2, contributions from this work and novel findings include:

- HAM2 negatively regulates vegetative growth (Chapter 3)
- HAM2 negatively regulates basal defence (Chapter 3)
- An *in silico* high-throughput screening approach to discover HAM2-specific inhibitors for agricultural applications (Chapter 4)

Overall, these results indicate that HAG1 and HAM2 have highly divergent roles even though they perform the same catalytic function of transferring acetyl moieties to histone lysine residues. The phenotypes of *ham2* and *hag1* mutant lines uncovered in this work are summarised in Figure 6.1.

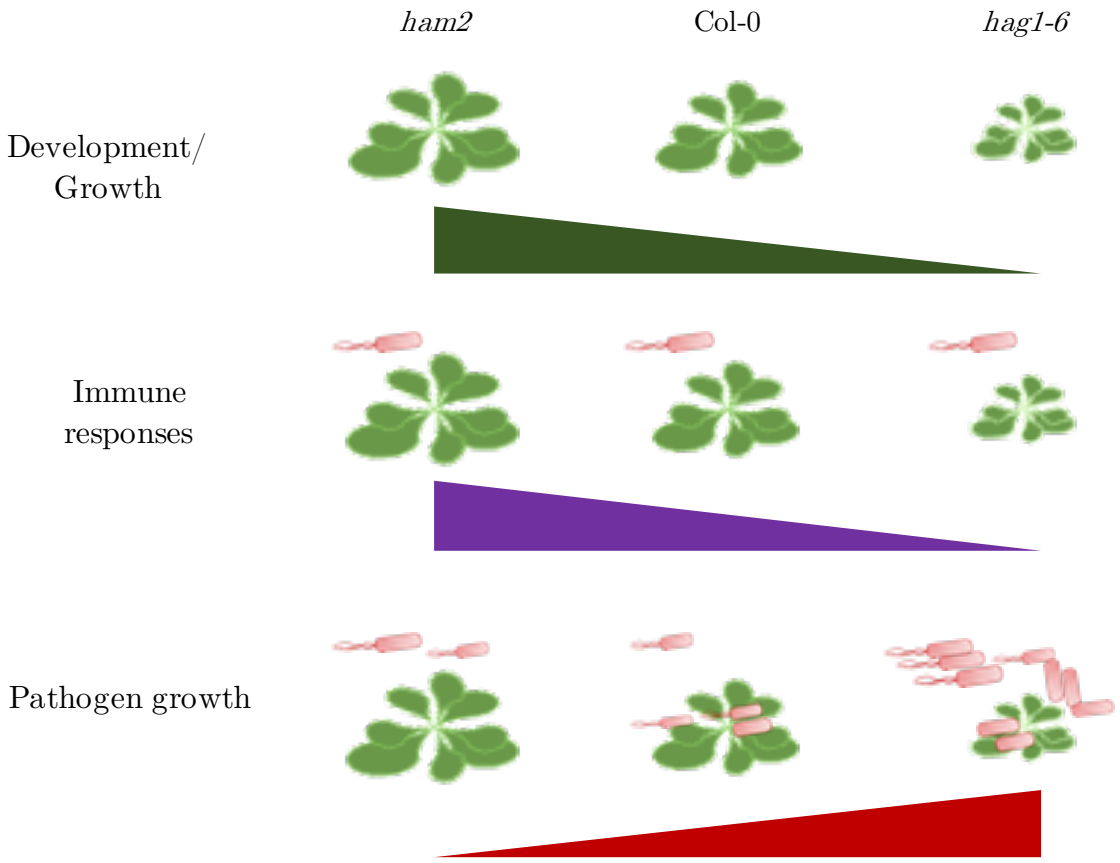


Figure 6.1: A summary of the *ham2* and *hag1* phenotypes presented in this thesis. *ham2* mutant lines show enhanced vegetative growth and enhanced resistance to *Pto* DC3000 (Chapter 3). The *hag1-6* mutant shows impaired vegetative growth and impaired resistance to *Pto* DC3000 (Chapter 5).

6.1 HAM2 moderates plant immunity and growth

In Chapter 3 of this thesis, several previously undescribed roles for the MYST-family HAT, HAM2, were uncovered. The finding that *ham2* mutant lines were more resistant to *Pto* DC3000 prompted the transcriptomic analysis of early PTI responses in the mutant. It was found that genes containing WRKY18-binding motifs were more highly expressed in the null mutant line *ham2-75* than in Col-0 after flg22 elicitation. This suggests that HAM2 acts as a transcriptional repressor to moderate defence gene expression. Immunity must be tightly regulated, since exacerbated defence responses leads to auto-immunity with developmental costs and reduced seed yield (Dietrich et al., 1994; Bowling et al., 1997). We therefore propose a working model,

presented in Figure 6.2, where HAM2 modulates the expression of defence-related genes upon activation of PTI to reduce potential development costs.

This work also presents the vegetative growth phenotypes of *ham2* mutants: under basal conditions, the *ham2* lines show enhanced leaf surface area and root length. A fundamental question that remains is that, if reduction of *HAM2* expression results in these beneficial phenotypes, why has *HAM2* not been lost or mutated during the course of evolution? Although beneficial phenotypes have been described in this work, several publications show that *ham2* mutants are negatively affected in gametophyte development (Latrasse et al., 2008) and UV-B dependent DNA damage repair (Campi et al., 2012). Therefore, HAM2 functions as a positive and negative regulator of pleitropic responses and abrogation of *HAM2* expression can indeed be deleterious. Realising this, the work presented here then moved towards identifying an inhibitor of HAM2 catalytic activity that would only be applied when required during the life cycle of the plant.

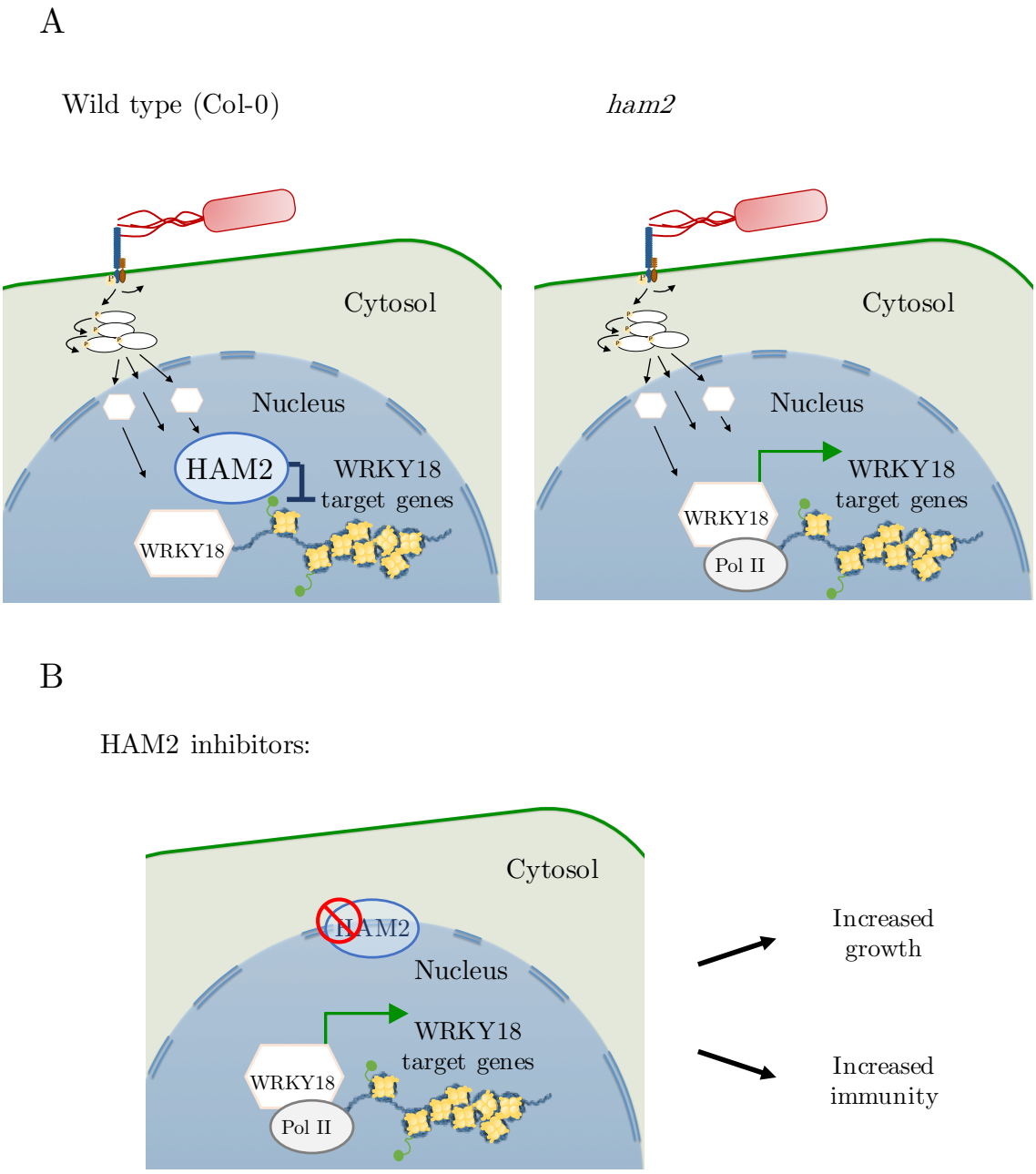


Figure 6.2: The role of HAM2 in basal defence responses. (A) During PTI, HAM2 acts as a transcriptional repressor of the expression of defence-related genes, such as WRKY18-motif containing genes, to tightly modulate early defence responses. In *ham2* lines, defence gene expression is increased leading to enhanced resistance to *Pto* DC3000. (B) Targeting HAM2 using chemical inhibitors to enhance defence responses in temporally controlled way.

6.2 Targeting chromatin modulators for crop improvement

In Chapter 4 of this thesis, the structures of *Arabidopsis* HAM2 and HAM2 homologues for economically important crop species were modelled. From the phenotypes of *Arabidopsis ham2* knock-down and knock-out lines presented in this thesis, and the transcriptomic differences between *ham2-75* and Col-0 during early PTI responses, it was clear that HAM2 acts as a modulator of expression at many genomic loci. From an agricultural perspective, we realised the potential advantages of manipulating HAM2 activity in a controlled manner, such that resistance to hemibiotrophic pathogens is enhanced and yield losses due to biotic stress is limited. By applying this control of HAM2 activity only when needed, the negative effects of *HAM2* mutation reported in the literature (Latrasse et al., 2008; Campi et al., 2012) can be limited. The *in silico* screen identified several candidate MYST-specific inhibitors from a library of compounds filtered for agrochemical likeness. Further experimental work will be performed to determine if the compounds specifically inhibit plant MYST proteins *in vivo*, and to further establish the kinetics of enzyme inhibition. Also, the compounds identified in this work could act as scaffolds for further development to optimise inhibitory activity.

In the field of crop biotechnology, there is an increasing interest in the potential for epigenetic variation to contribute to improvement of crop performance (recently reviewed by Springer and Schmitz (2017)). It is thought that diversity in morphological traits could arise from cryptic allelic variation (i.e., allelic differences that are only observed when the epigenetic state of a loci is modified) (Springer, 2013). Indeed, one strategy to generate "epimutants" to uncover desired traits has been to apply chemical treatments that affect genome-wide DNA methylation patterns (Akimoto et al., 2007). A draw-back of the epimutant approach is that they may not be as stable as genetic mutants, and it may be difficult to maintain the desired epigenetic state from one generation to the next. In the current study, it was discovered that altering expression of *HAM2* led to agriculturally attractive phenotypes as HAM2 is likely to function as a transcriptional regulator at many different genomic loci. The work initiated in the present study will lead to the identification and optimisation of an agrochemical with specific inhibitory kinetics. In this way, the beneficial morphological traits uncovered by altering *HAM2* expression will not depend on generation to generation transmission; instead the grower will be able to

induce these epigenetic alterations when desired in a condition and time-dependent manner.

6.3 HAG1 plays a central role in ETI

To date, there have been no publications describing a role for *Arabidopsis* HATs in effector-triggered immunity. As described in Section 1.3.3, the transcriptional changes induced by PTI and ETI have been found to overlap extensively (Navarro et al., 2004), and there are many similarities between PTI and ETI. In fact, it is thought that the differences between PTI and ETI do not occur from alternate signalling machinery, but instead differences arise in duration and amplitude of responses (Tsuda and Katagiri, 2010). Given the key role that HATs play in transcriptional regulation, and that we have identified *hat* mutants with altered PTI immune responses (Piquerez et al., in preparation), we hypothesised that HATs contribute to effector-triggered responses. A screen was performed to assess AvrRpt2-induced HR in *Arabidopsis hat* mutants and, as well as *hac4*, *hac5* and *ham1*, *hag1* was found to have defective responses to AvrRpt2.

Further bacterial growth and ion leakage assays demonstrated that HAG1 was also a positive regulator of AvrRpm1, AvrPphB and AvrRps4-induced immunity. Transcriptomic analysis following infection with *Pto* (*avrRpt2*) revealed the extent of defective responses in *hag1-6*: approximately 2000 genes which were differentially expressed in Col-0 in response to *Pto* (*avrRpt2*) were not in this mutant. Based on these findings, a working model of HAG1 function during ETI responses is presented in Figure 6.3. Since HAG1 positively regulates AvrRpt2, AvrRpm1, AvrPphB and AvrRps4-induced immunity, HAG1 is likely to function at a downstream point when effector-specific signals are converged. Also, since the expression of so many genes were not up-regulated in response to *Pto* (*avrRpt2*) in *hag1-6*, HAG1 most likely acts as a transcriptional activator for ETI-induced genes. Further experiments to clarify how HAG1 is targeted to alternate loci following induction of ETI were described in Section 5.5. Another approach to identify transcription factor binding partners of HAG1, and whether these partners change in a condition-dependent manner, would be to perform mass spectrometry on Col-0 and *hag1-6* plants following infection with *Pto* expressing different Avr proteins.

As mentioned previously, it is thought that the differences between PTI and ETI do

not occur from alternate signalling machinery (Tsuda and Katagiri, 2010); based on the findings presented here, HAG1 is a component of the machinery that functions to positively regulate both PTI and ETI. Also, Kong et al. (2017) recently published that PsAvh23, an effector deployed by *Phytophthora sojae*, interferes with the catalytic activity of soybean HAG1 leading to increased pathogen colonisation. It is known that effectors target both host PTI and ETI responses; therefore, the pathogen *Phytophthora sojae* could be interfering with soybean HAG1 to attenuate PTI and ETI defences to promote successful disease progression.

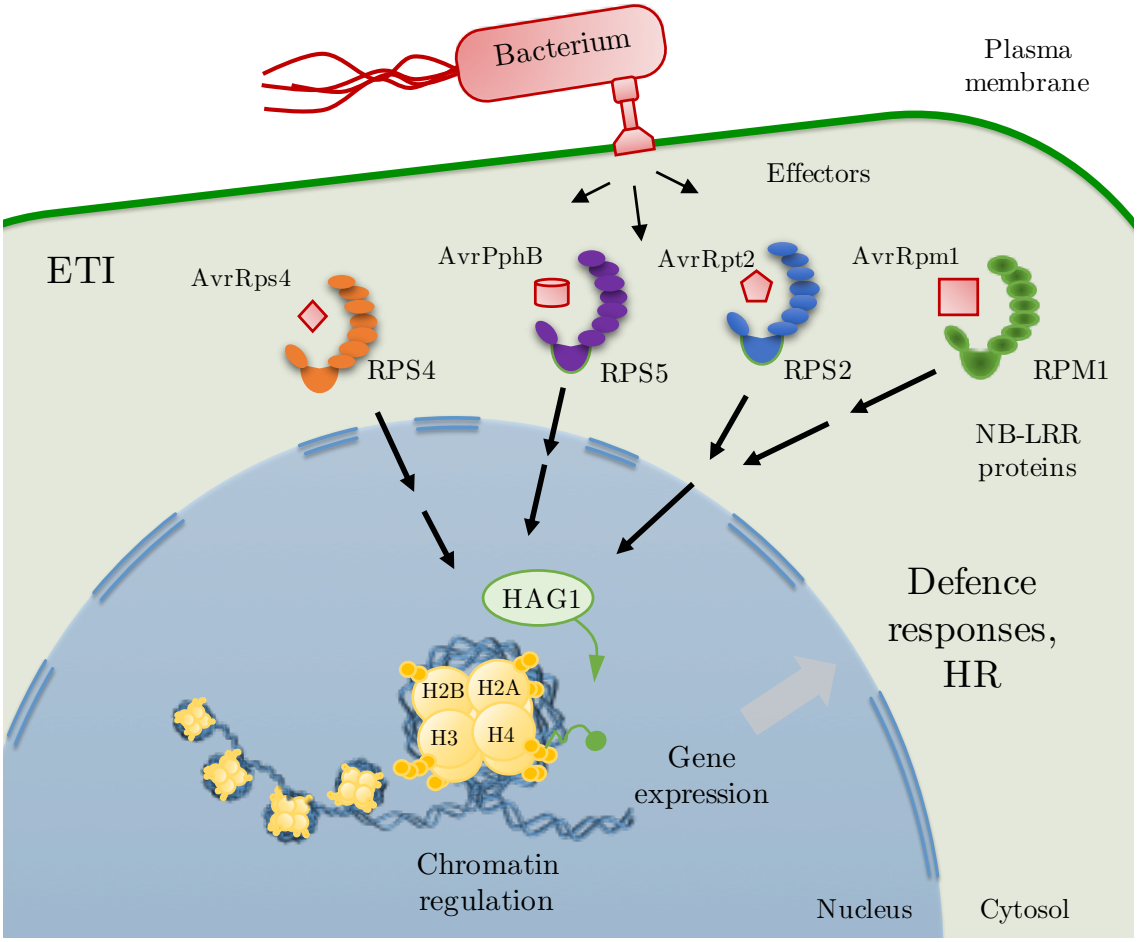


Figure 6.3: HAG1 plays a central role in ETI following Avr protein perception. Following host detection of a pathogenic effector, HAG1 strongly modulates the expression of ETI-induced genes to contribute to the HR in order to limit pathogen growth.

6.4 General outlook

The work presented in this thesis improves our molecular understanding of how plants establish effective defences against potential pathogens. Whilst being scientifically intriguing, the over-arching motivation of plant research is to translate knowledge gained using model organisms into advances with real-world benefits. In this work, desirable traits seen in *Arabidopsis* mutant lines were identified and it is hoped that, through this work, these traits will be replicated in economically important crops to contribute to food security. This work also highlights the prospective benefits of incorporating epigenetic research into plant breeding programmes, as discussed in Section 6.2.

Also highlighted in this work is the potential when multiple disciplines and approaches are harnessed to tackle research questions. Transcriptomic and bioinformatic technologies are now regularly used to decipher how plants respond to various stresses. These commonly used techniques were successfully adopted in this work to discover how HATs regulate gene expression when under biotic stress. Also, in an approach more commonly used in the pharmaceutical industry, *in silico* modelling and screening techniques were adopted in order to identify a novel agrochemical. By taking inspiration from other sectors of scientific research, innovative solutions and technologies can be developed to contribute to solving the global issue of food security. It should be noted that research and development efforts should be complemented with improved communication with the public in order to improve societal perception and trust in science. In this way, legislative bodies and the general public may become more accepting of transgenics and other technologies, as uptake of technologies ultimately depends upon this.

Appendix A

HAT T-DNA insertion mutant lines and primers for genotyping.

Arabidopsis line	SALK line	Forward primer (5' to 3')	Reverse primer (5' to 3')
LBb1.3 for SALK line genotyping			ATTTTGCCGATTTCGGAAC
hag1-6	SALK_150784	TTGCACAAAATGCTATTTCCC	CTCCAACGATGAACTCGAGAG
hag2-99	SALK_051832	CCGTATAAAACCAATCCAACG	TTGAATTGGTCTTGCATTTGG
haf2-29	SALK_110029	TGGCTAGCTCATTGTCATGTG	GGGATTCATGTATCATGGGTG
ham1-50	SALK_103726	AGAATCAGCCACTTCAACACG	GATTCTGAATTCGTGAGAGCG
ham1-96	SALK_027726	ATGGTGTGCGAATCTATGACC	ACGGAGAGGAAAGCTCAAGAC
ham2-71	SALK_012086	CCAATTCCAATGATCCAATTG	TTCACTCATGGATACTTCCGC
ham2-75	SALK_106046	GTCGAAGAAGAGGAAAATGGG	CATATGCCTTTGAAGCTGCTC
hac2-84	SALK_049434	ATCACATTTCATCGGCTCAAC	TTTCTTCTTCGTCCTGTCTGC
hac4-15	SALK_051750	CTAGGAGGCTTTGGTCACATG	CTTTGACCAAAAGAAGCAACG
hac4-21	SALK_045791	TCTTCATGCGTCCCATAGTTC	CATGTGTGACTGATCAGGTGG
hac5-78	SALK_024278	GATTTTTCCGGCTTTGATCTC	CAATGACAACCTGTGCAACAC
hac12-04	SALK_071102	TTGTCTATATCCCAACTGCCG	ATCCCCCAAAGGGATATAAGC

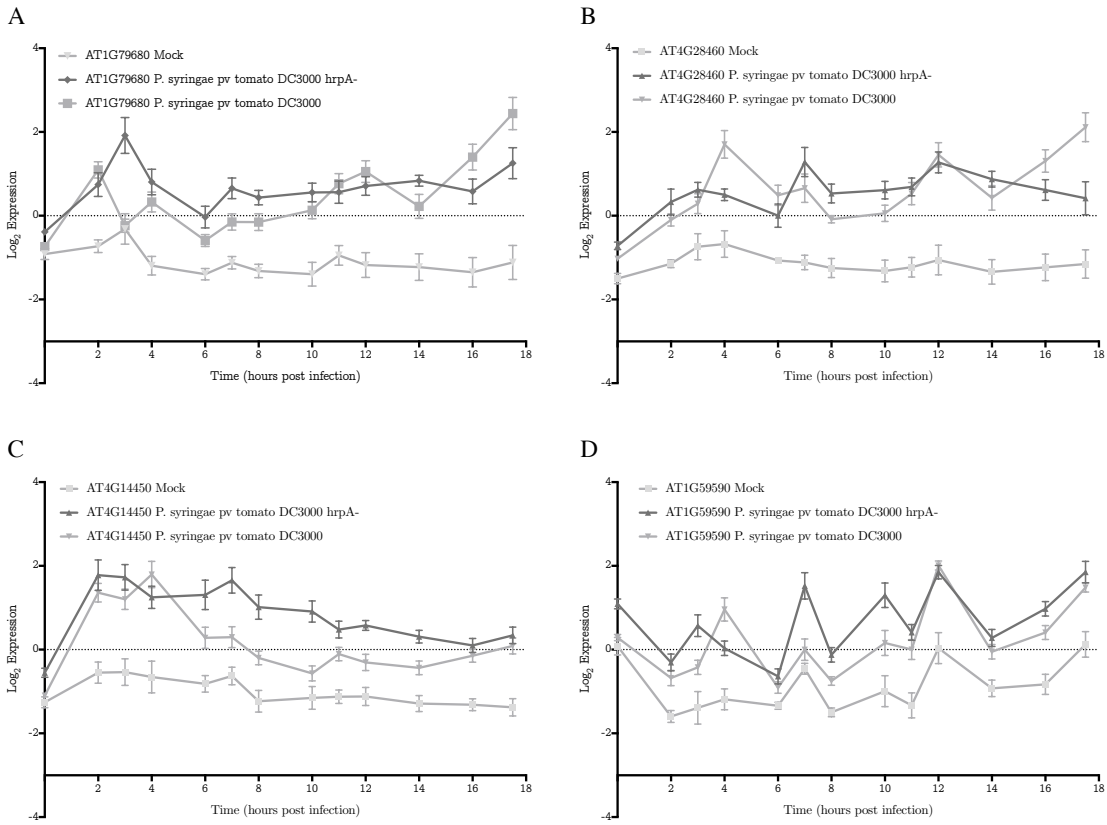
Primers for qPCR

Name	Use	Sequence (5' to 3')
qHAM2-F	Forward qPCR primer for HAM2	GTGATGTGTCGGTGGAGAGA
qHAM2-R	Reverse qPCR primer for HAM2	GCTTGTTACCTTGTCTTCCAC
qUbox-F	Forward housekeeping gene (AT5G15400)	TGCGCTGCCAGATAATACACTATT
qUbox-R	Reverse housekeeping gene (AT5G15400)	TGCTGCCCAACATCAGGTT
qAIG1-F	Forward qPCR primer of AIG1 (At1g33960)	AAGGGCATTCAGAAGAGCAGCT
qAIG1-R	Reverse qPCR primer of AIG1 (At1g33960)	CTCCTGCGCCTCCATAGCTATT
qAIG2-F	Forward qPCR primer of AIG2 (At3g28930)	CATGGTTTCCGCTCAACTCCAC
qAIG2-R	Reverse qPCR primer of AIG2 (At3g28930)	GCCTCTTCCATTCCTCGAAATCC
qWRKY46-F		AACCGACCAAGTCCGAAGAAGT
qWRKY46-R		CGGAATTCTCAACAGCAGCAGG

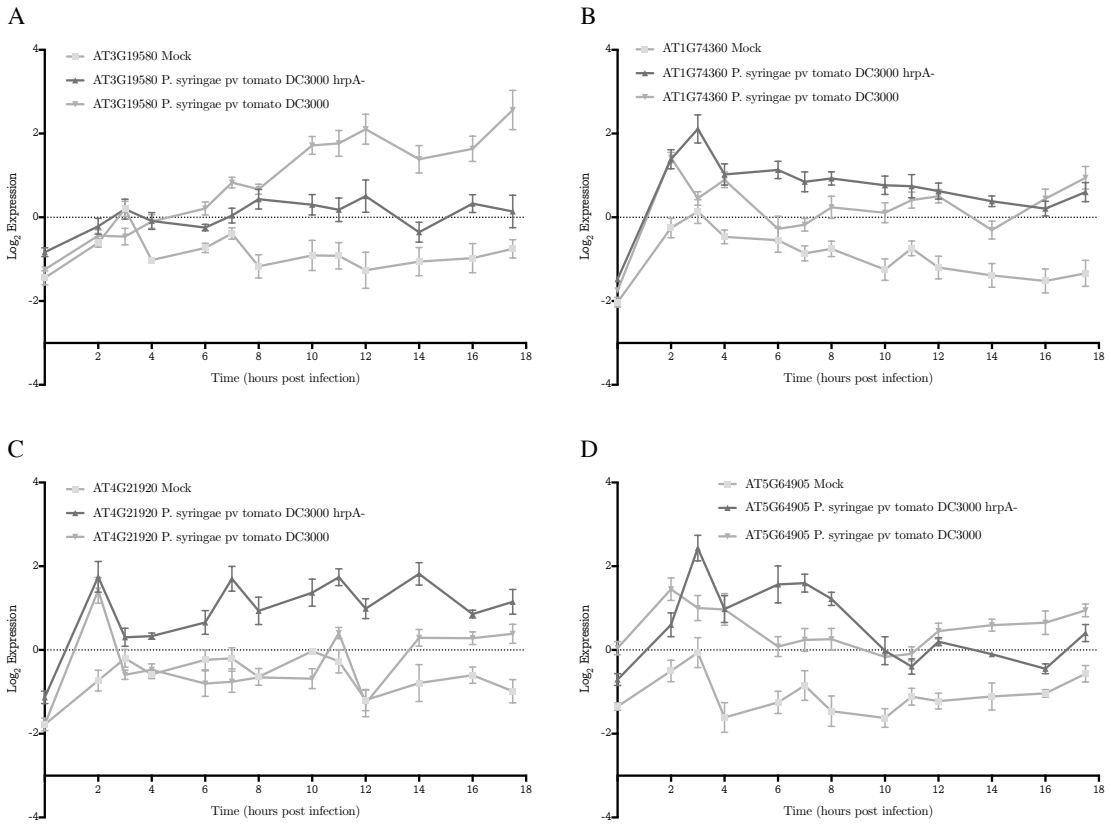
Primers for gateway cloning

Name	Use	Sequence (5' to 3')
AvrRpt2-F	Clone AvrRpt2 in pDONR-Zeo (Start)	AAAAAGCAGGCTCCACCATGAAAATTGCTCCAG
AvrRpt2-R	Clone AvrRpt2 in pDONR-Zeo (Stop)	AGAAAGCTGGGTCGCGGTAGAGCATT
AvrRpt2-FS2	AvrRpt2 sequencing primer (forward)	CACCCACTCAGGCAAAG
AvrRpt2-FS1	AvrRpt2 sequencing primer (forward) (pBAV)	GCACTGTTGTATAAGCACG
AvrPphB-F	Clone AvrPphB in pDONR-Zeo (Start)	AAAAAGCAGGCTCCACCATGAAAATAGGTA
AvrPphB-R	Clone AvrPphB in pDONR-Zeo (Stop)	AGAAAGCTGGGTCCGAAACTCTAAACTC
AvrPphB-FS2	AvrPphB sequencing primer (forward)	GGCCGTTTTGCACAATCA
AvrPphB-FS1	AvrPphB sequencing primer (forward) (pBAV)	AATCACGCATGGACCATC
attB1	Gateway cloning: full attB sites (forward)	GGGGACAAGTTTGTACAAAAAAGCAGGCT
attB2	Gateway cloning: full attB sites (reverse)	GGGGACCACTTTGTACAAGAAAGCTGGGT

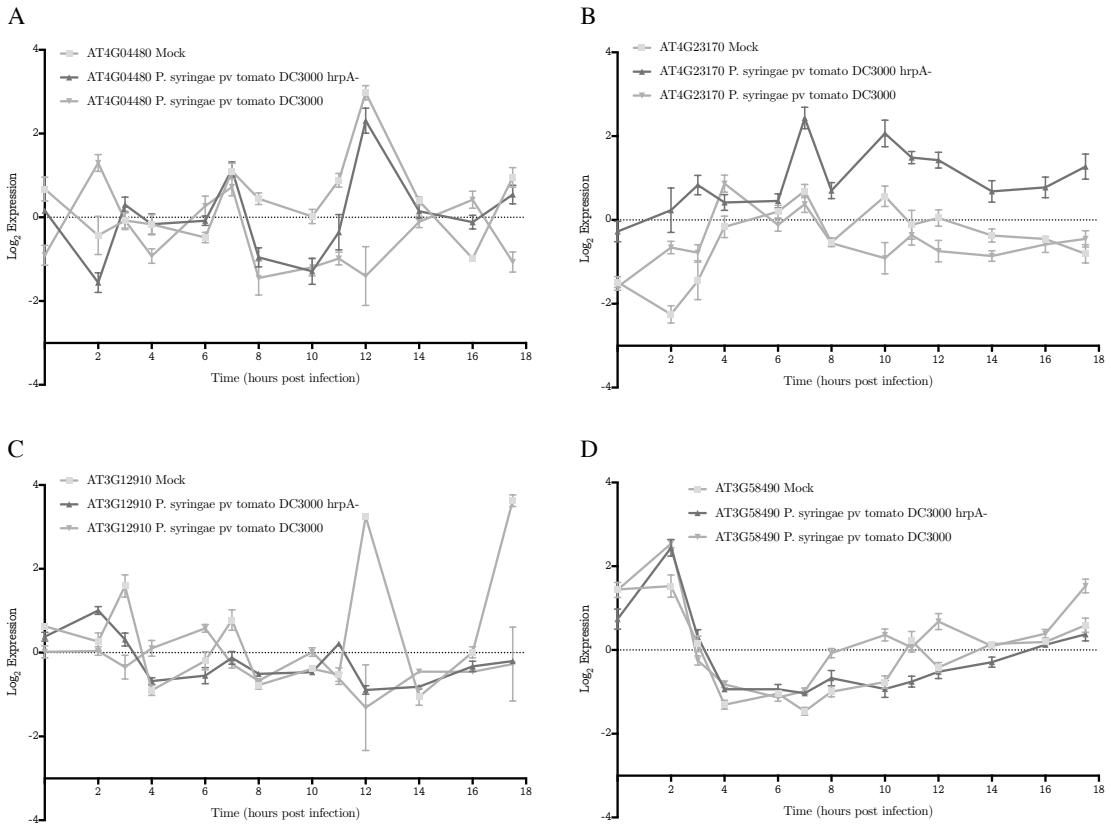
Appendix B



WRKY18 binding motif containing genes in response to infection with *Pto* DC3000 and *Pto* DC3000*hrpA*- (Lewis et al. (2015), GEO accession number: GSE56094).

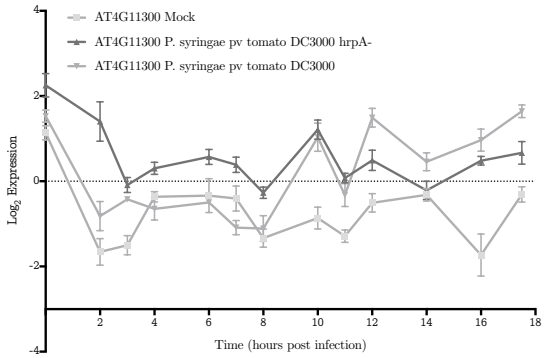


WRKY18 binding motif containing genes in response to infection with *Pto* DC3000 and *Pto* DC3000*hrpA*- (Lewis et al. (2015), GEO accession number: GSE56094).



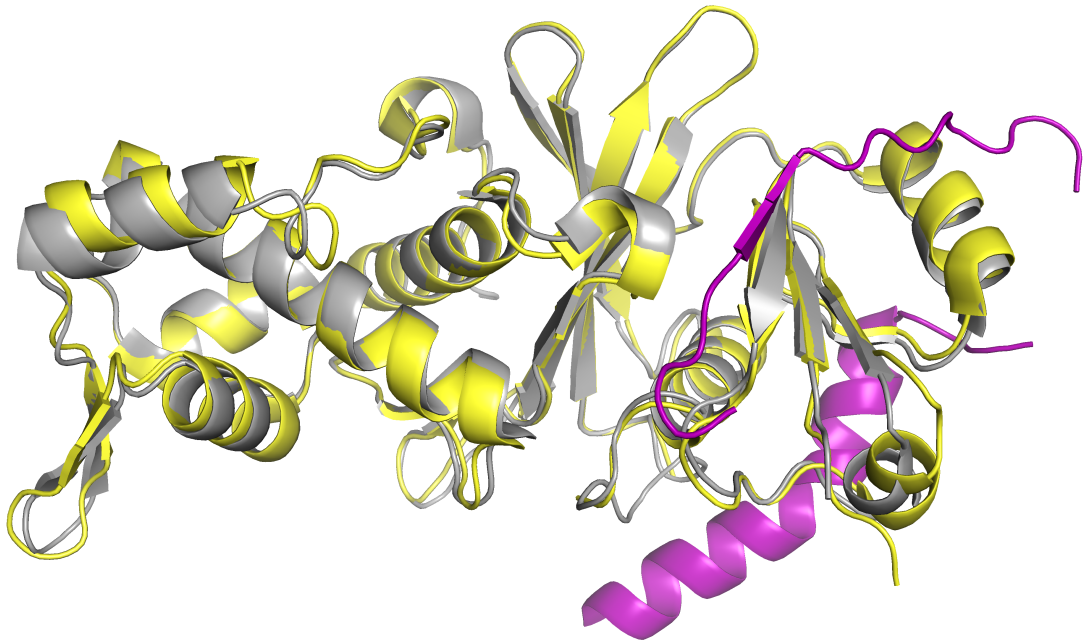
WRKY18 binding motif containing genes in response to infection with *Pto* DC3000 and *Pto* DC3000*hrpA*- (Lewis et al. (2015), GEO accession number: GSE56094).

A

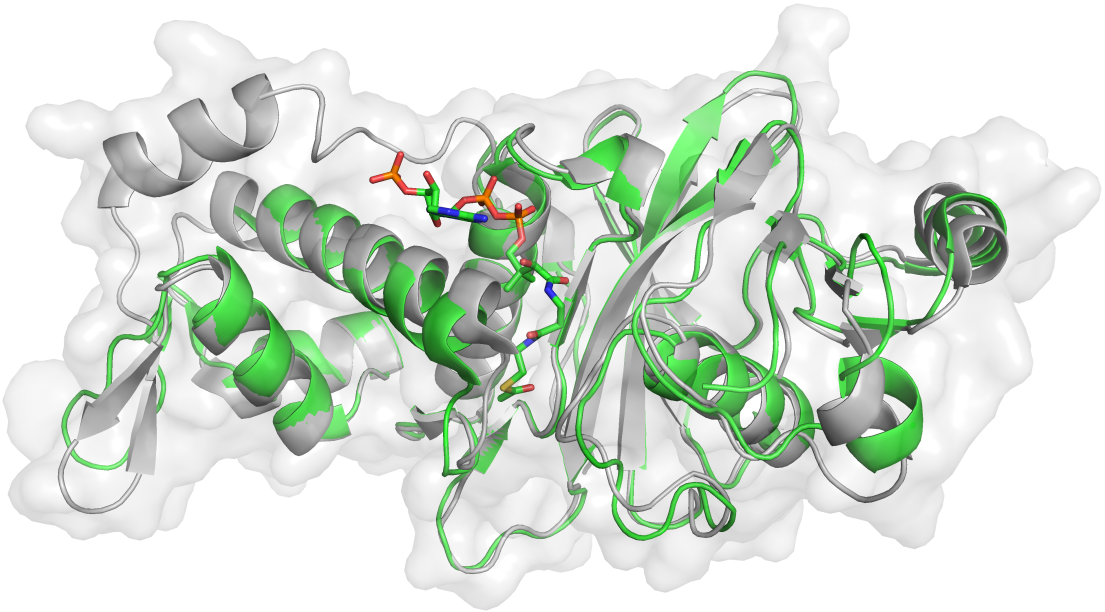


WRKY18 binding motif containing genes in response to infection with *Pto* DC3000 and *Pto* DC3000*hrpA*- (Lewis et al. (2015), GEO accession number: GSE56094).

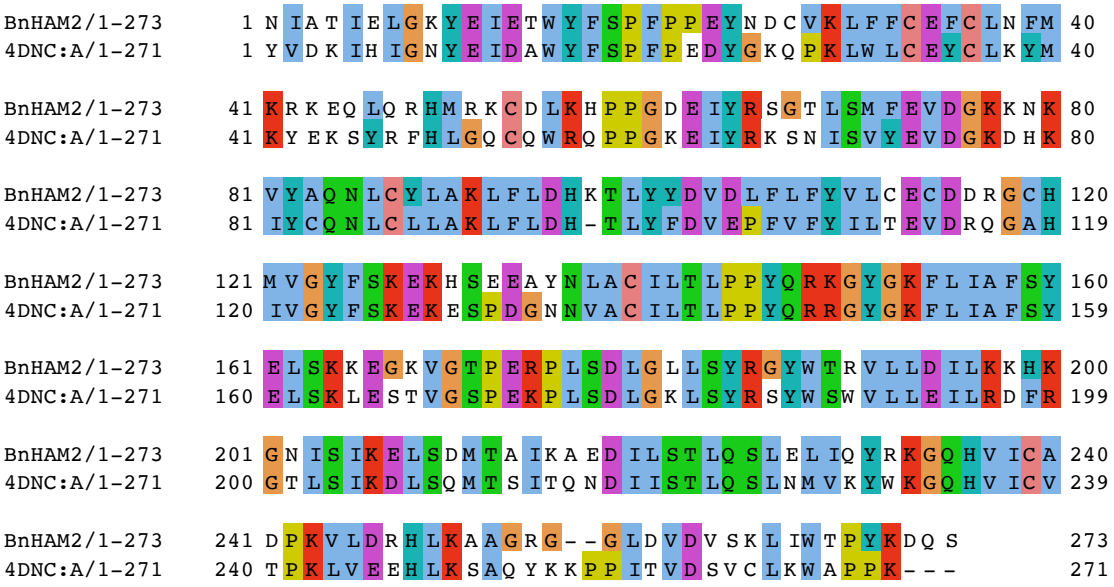
Appendix C



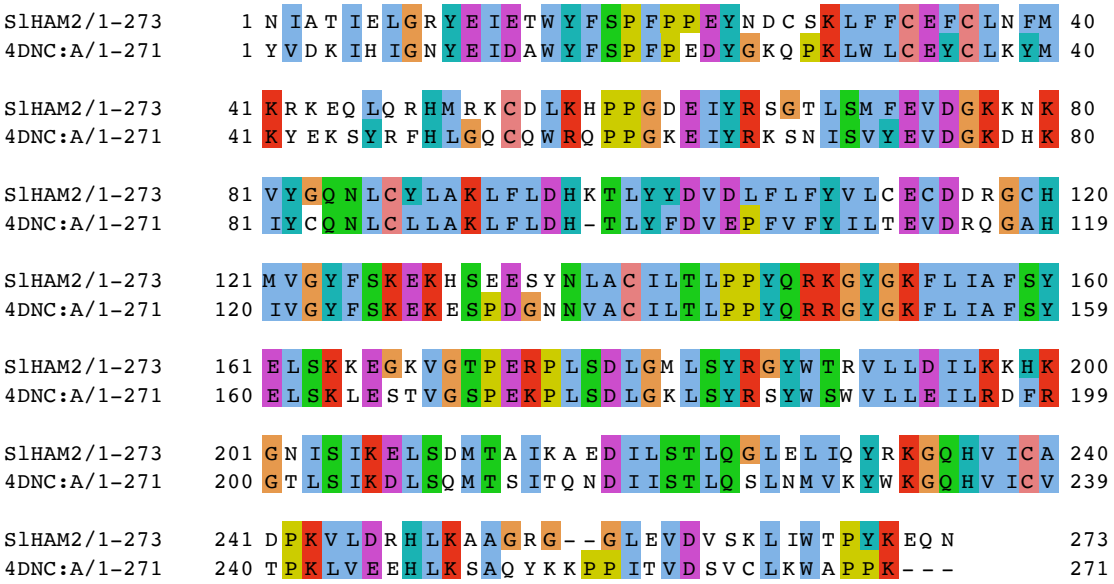
Structural comparison of PDB ID 4DNC and 5J8F. MSL1 (pink) does not cause a conformational change in HsKAT8. This was assessed by aligning 4DNC (grey) with 5J8F (HsKAT8 not co-crystallised with a ligand, yellow): the two structures align with RMSD = 0.624 Å between Cα atoms and MSL1 does not cause a conformational change in the backbone structure of 4DNC.



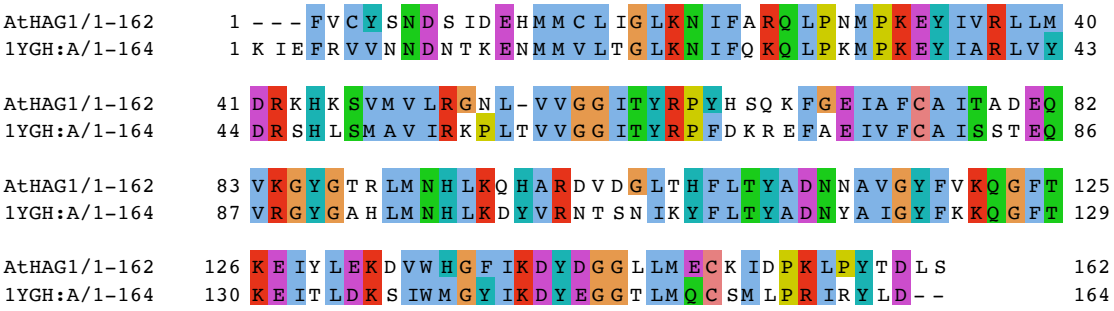
Structural comparison of PDB ID 4DNC and 2OU2. Superimposition of 4DNC (HsKAT8 co-crystallised without Ac-CoA, grey) and 2OU2 (HsKAT8 co-crystallised with Ac-CoA, green) indicates that a conformational change is induced at α-helix 4 as a result of acetyl-CoA binding. Protein structures are presented in cartoon form; acetyl-CoA structure is presented in stick form.



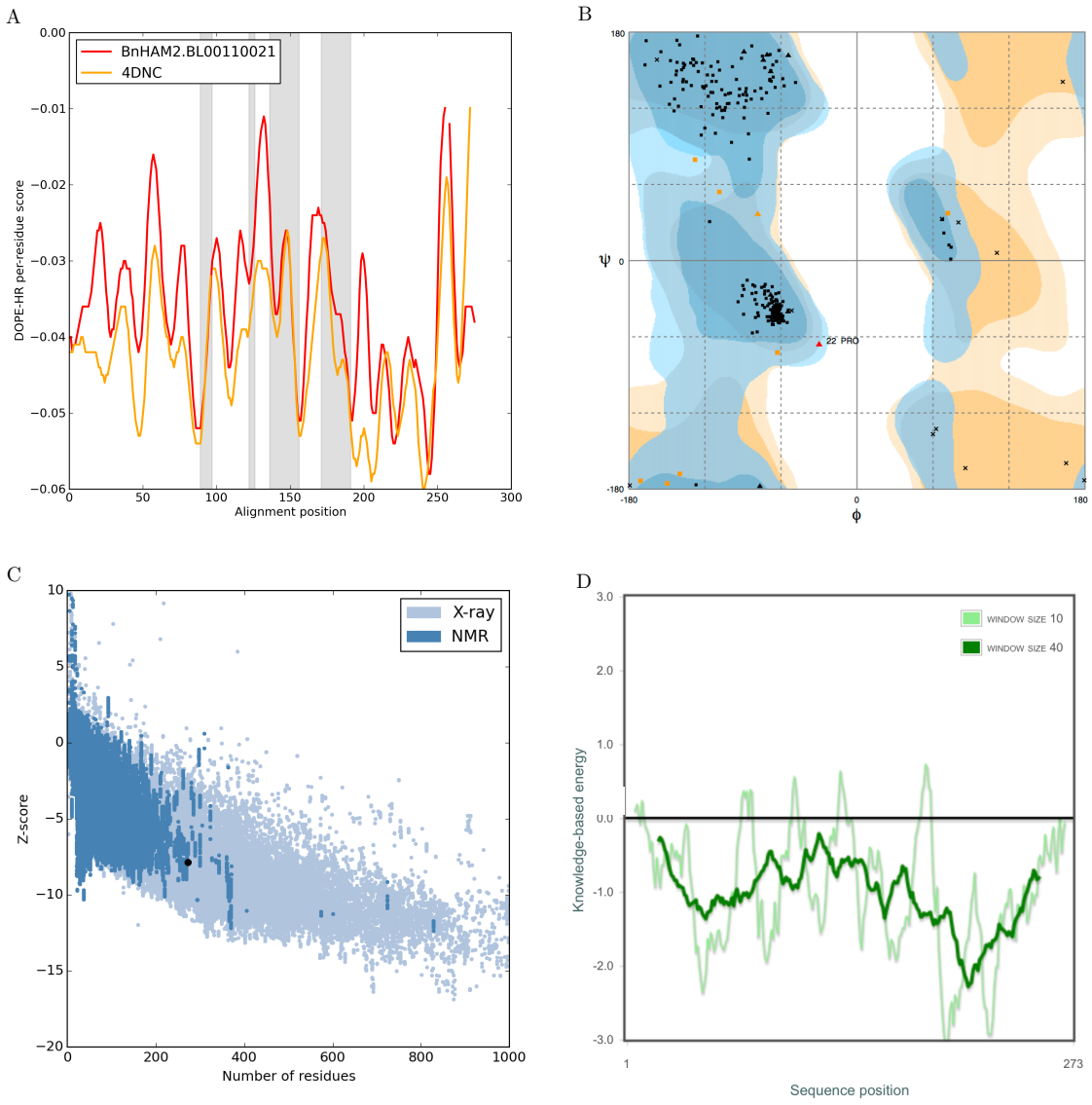
Sequence alignment of the HAT domains of HsKAT8 (PDB: 4DNC) and BnHAM2. The same colouring for aligned residues indicates conservation of amino acid chemical properties (ClustalX colouring, blue: hydrophobic, red: positively charged, magenta: negative charged, green: polar, cyan: aromatic, pink: cysteine, yellow: proline, orange: glycine).



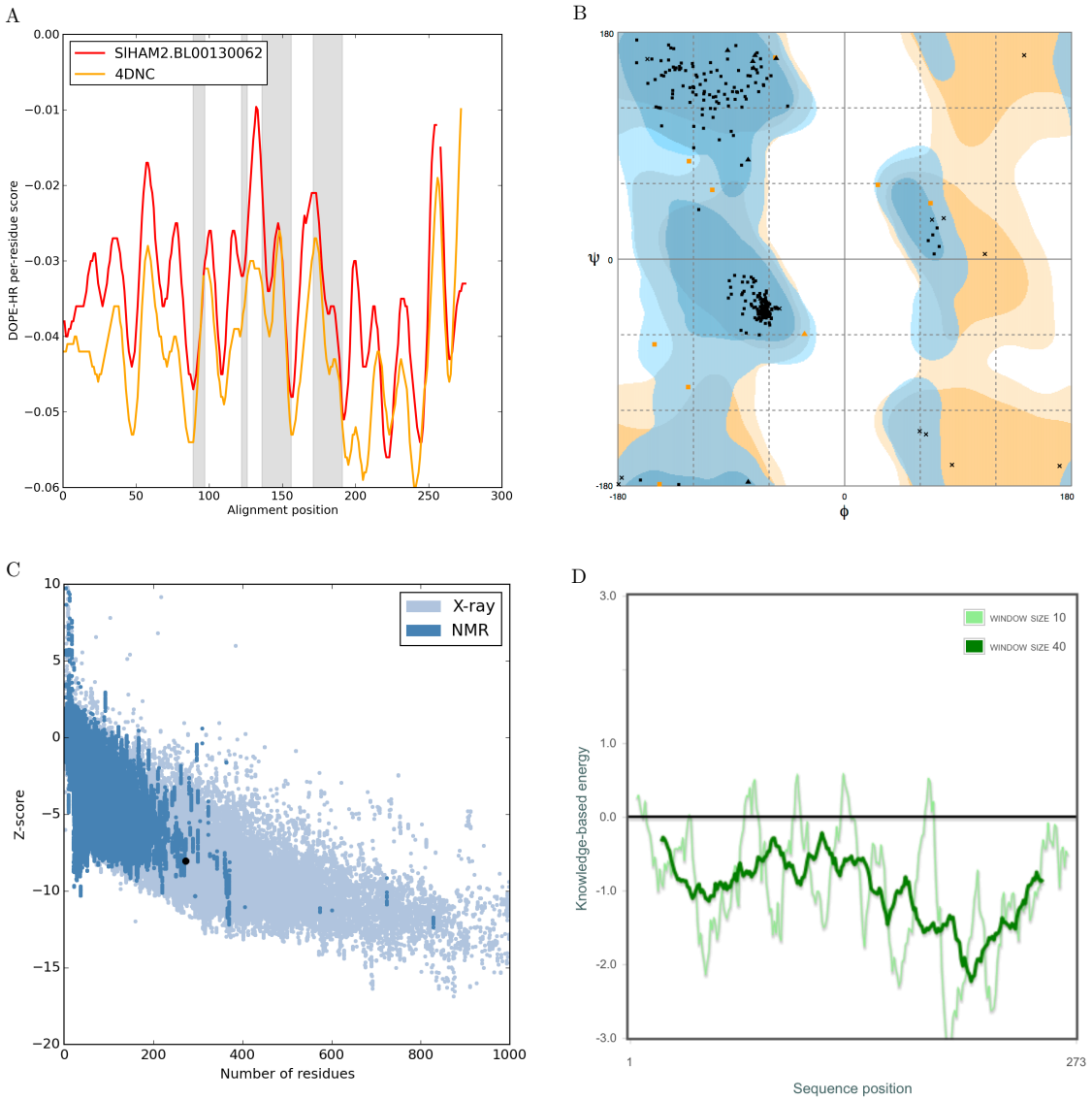
Sequence alignment of the HAT domains of HsKAT8 (PDB: 4DNC) and SlHAM2. The same colouring for aligned residues indicates conservation of amino acid chemical properties (ClustalX colouring, blue: hydrophobic, red: positively charged, magenta: negative charged, green: polar, cyan: aromatic, pink: cysteine, yellow: proline, orange: glycine).



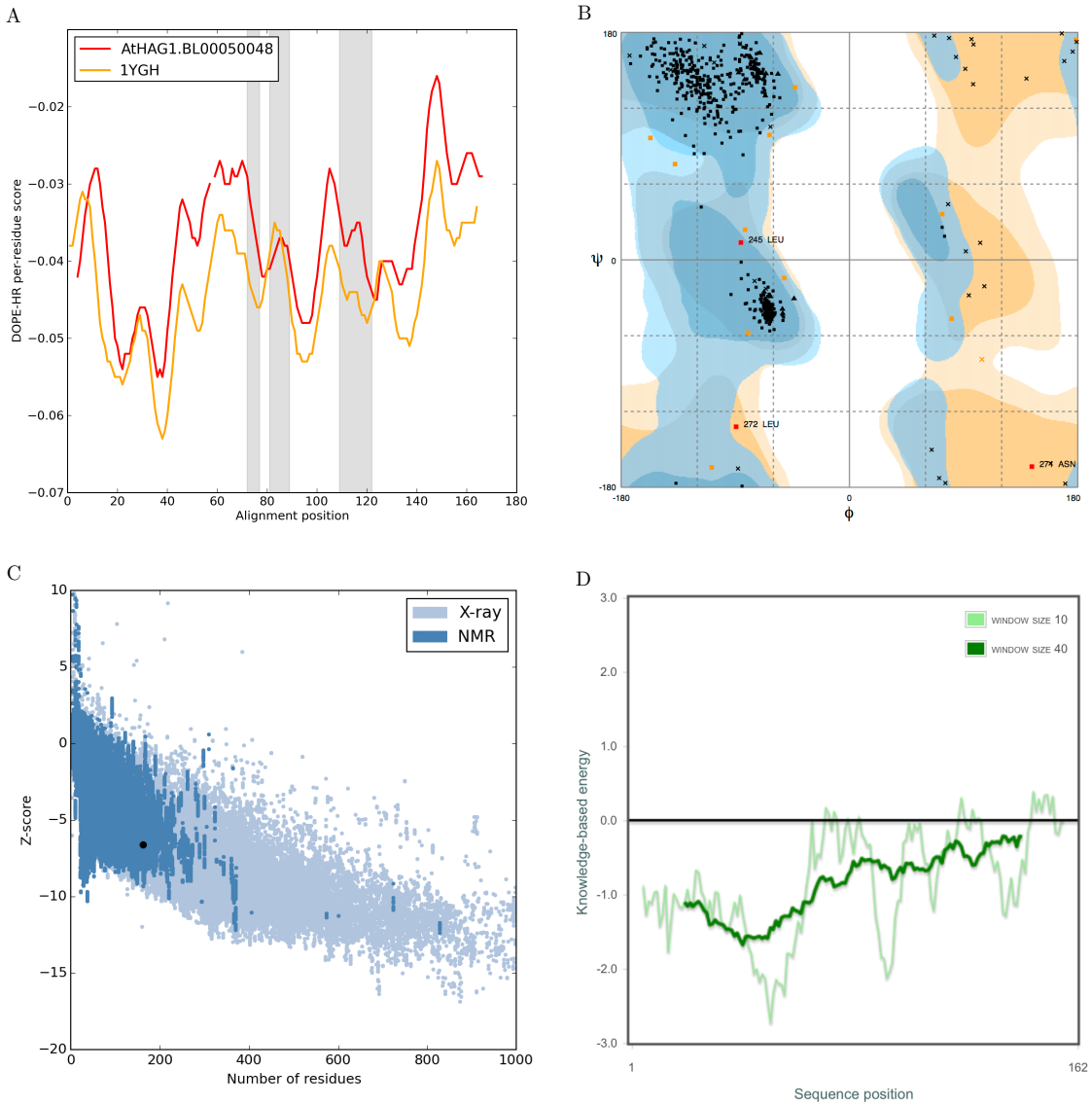
Sequence alignment of the HAT domains of ScGCN5 (PDB: 1YGH) and AtHAG1. The same colouring for aligned residues indicates conservation of amino acid chemical properties (ClustalX colouring, blue: hydrophobic, red: positively charged, magenta: negative charged, green: polar, cyan: aromatic, pink: cysteine, yellow: proline, orange: glycine).



Evaluation of BnHAM2 homology model quality. (A) DOPE per-residue score of the template (4DNC, orange) and the target (BnHAM2, red). Residues forming the Ac-CoA binding pocket are shaded in grey. (B) Ramachandran plot analysis. Areas shaded in dark blue, light blue and yellow refer to the "core", "allowed" and "generously allowed" regions respectively. Black dots indicate residue lies within "core" region, orange dots in the "allowed" region and red dots in the "disallowed" region. (C) ProSA z-score of BnHAM2 model (black dot) plotted with Z-scores of all experimentally determined structures currently in PDB (dark blue dots: by NMR, light blue dots: by X-ray). (D) ProSA per-residue interaction energy of the model. Light green line: 10-residue fragment window; dark green line: 40-residue fragment window.



Evaluation of SlHAM2 homology model quality. (A) DOPE per-residue score of the template (4DNC, orange) and the target (SlHAM2, red). Residues forming the Ac-CoA binding pocket are shaded in grey. (B) Ramachandran plot analysis. Areas shaded in dark blue, light blue and yellow refer to the "core", "allowed" and "generously allowed" regions respectively. Black dots indicate residue lies within "core" region, orange dots in the "allowed" region and red dots in the "disallowed" region. (C) ProSA z-score of SlHAM2 model (black dot) plotted with Z-scores of all experimentally determined structures currently in PDB (dark blue dots: by NMR, light blue dots: by X-ray). (D) ProSA per-residue interaction energy of the model. Light green line: 10-residue fragment window; dark green line: 40-residue fragment window.



Evaluation of AtHAG1 homology model quality. (A) DOPE per-residue score of the template (4DNC, orange) and the target (AtHAG1, red). Residues forming the Ac-CoA binding pocket are shaded in grey. (B) Ramachandran plot analysis. Areas shaded in dark blue, light blue and yellow refer to the "core", "allowed" and "generously allowed" regions respectively. Black dots indicate residue lies within "core" region, orange dots in the "allowed" region and red dots in the "disallowed" region. (C) ProSA z-score of AtHAG1 model (black dot) plotted with Z-scores of all experimentally determined structures currently in PDB (dark blue dots: by NMR, light blue dots: by X-ray). (D) ProSA per-residue interaction energy of the model. Light green line: 10-residue fragment window; dark green line: 40-residue fragment window.

Top overlap scorers (1:40). The overlap is a value representing the fraction of AtHAM2 atoms that interact with Ac-CoA that also interact with the test compound.

Name	**Overlap**	**Overlap dissim.**	**RMSD**
7_12927	0.723404	-0.411765	0.163159
23_620	0.680851	0.28125	0.159546
46_807	0.680851	0.0625	0.176233
21_18836	0.659574	-0.16129	0.189822
20_10679	0.638298	0.3	0.184776
9_17598	0.638298	0.3	0.202939
23_7782	0.638298	0.2	0.192543
23_10476	0.638298	0.2	0.197284
40_14139	0.638298	0.2	0.177016
33_19987	0.638298	0.133333	0.190849
44_2658	0.638298	0.0666667	0.184054
33_2721	0.638298	-0.0666667	0.16162
7_7589	0.638298	-0.1	0.189995
48_9501	0.638298	-0.366667	0.175448
22_19692	0.617021	0.413793	0.215417
21_7609	0.617021	0.275862	0.199073
24_3535	0.617021	0.206897	0.190876
20_14474	0.617021	0.172414	0.17505
5_5427	0.617021	0.172414	0.180803
31_7338	0.617021	0.137931	0.164562
28_16472	0.617021	0.103448	0.180695
34_7858	0.617021	0.0689655	0.186368
45_10758	0.617021	0.0689655	0.157123
23_7228	0.617021	0.0344828	0.167132
43_17189	0.617021	0.0344828	0.209307
30_13564	0.617021	-0.0344828	0.198737
47_16564	0.617021	-0.0689655	0.207616
5_7424	0.617021	-0.137931	0.17801
9_16610	0.617021	-0.172414	0.192974
23_9197	0.617021	-0.172414	0.15806
12_15636	0.617021	-0.206897	0.184468
7_11693	0.617021	-0.206897	0.21297
30_18622	0.617021	-0.241379	0.191116
34_14437	0.617021	-0.275862	0.180575
25_5311	0.617021	-0.310345	0.17371
11_2838	0.595745	0.571429	0.160256
26_13420	0.595745	0.285714	0.166208
19_16660	0.595745	0.25	0.179031
49_16984	0.595745	0.25	0.167472
10_16175	0.595745	0.214286	0.182838

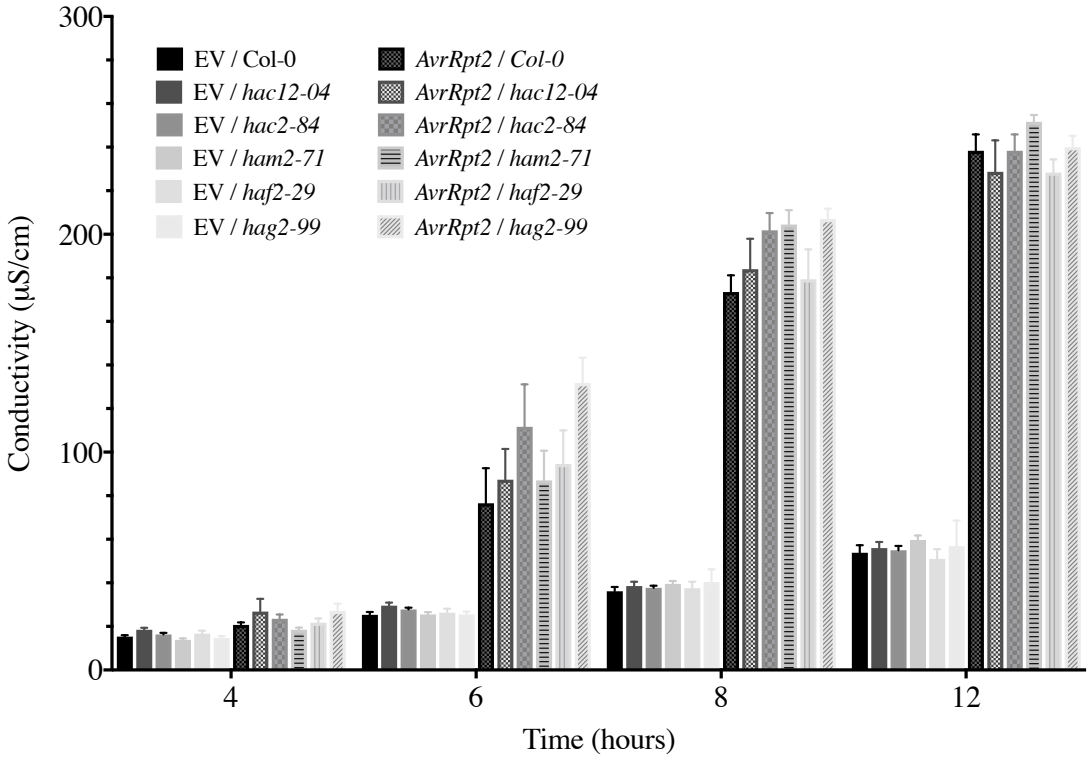
Top overlap scorers (41:80). The overlap is a value representing the fraction of AtHAM2 atoms that interact with Ac-CoA that also interact with the test compound.

Name	**Overlap**	**Overlap dissim.**	**RMSD**
9_3152	0.595745	0.178571	0.14846
39_7439	0.595745	0.178571	0.179947
40_19726	0.595745	0.142857	0.156915
33_8924	0.595745	1.43E-01	0.175289
48_12636	0.595745	0.107143	0.144526
19_2993	0.595745	0.107143	0.207922
48_9578	0.595745	7.14E-02	0.179676
40_16112	0.595745	0.0714286	0.141093
11_15721	0.595745	0.0714286	0.170853
11_15722	0.595745	0.0714286	0.196405
29_8727	0.595745	0.0357143	0.150308
37_604	0.595745	0.0357143	0.168272
18_755	0.595745	0.0357143	0.211424
9_15766	0.595745	2.78E-17	0.200055
35_5376	0.595745	2.78E-17	0.185847
36_4952	0.595745	0	0.176155
25_11765	0.595745	-1.39E-17	0.194876
17_9978	0.595745	-0.0357143	0.183632
28_5106	0.595745	-0.0357143	0.146598
23_17001	0.595745	-0.0714286	0.15669
24_9714	0.595745	-0.0714286	0.182441
35_2272	0.595745	-0.107143	0.147561
48_10261	0.595745	-0.142857	0.143234
13_2579	0.595745	-0.142857	0.185945
44_15335	0.595745	-0.142857	0.166327
22_9573	0.595745	-0.142857	0.175888
29_13079	0.595745	-0.214286	0.193878
15_6838	0.595745	-0.25	0.221368
37_7730	0.595745	-0.285714	0.198356
43_5138	0.574468	0.407407	0.198617
48_7499	0.574468	0.37037	0.182938
12_2959	0.574468	0.37037	0.21907
8_12725	0.574468	0.333333	0.163922
23_10388	0.574468	0.333333	0.164227
47_12504	0.574468	0.296296	0.208164
45_2779	0.574468	0.259259	0.15069
7_8653	0.574468	0.259259	0.182479
37_13924	0.574468	0.259259	0.15142
41_3856	0.574468	0.259259	0.231115
27_3767	0.574468	0.222222	0.200493

Top overlap scorers (81:100). The overlap is a value representing the fraction of AtHAM2 atoms that interact with Ac-CoA that also interact with the test compound.

Name	**Overlap**	**Overlap dissim.**	**RMSD**
9_6345	0.574468	0.185185	0.174514
47_584	0.574468	0.185185	0.200139
39_10928	0.574468	0.185185	0.187816
19_9788	0.574468	0.185185	0.163452
22_1830	0.574468	0.148148	0.19714
24_2941	0.574468	0.148148	0.197839
11_7905	0.574468	0.148148	0.180596
30_17867	0.574468	0.148148	0.169251
39_4780	0.574468	0.148148	0.194911
12_16905	0.574468	0.148148	0.211496
40_18712	0.574468	0.148148	0.164454
12_17196	0.574468	0.148148	0.153865
31_18486	0.574468	0.111111	0.190126
10_3576	0.574468	0.111111	0.186808
49_5258	0.574468	0.111111	0.207371
38_11224	0.574468	0.0740741	0.180318
8_10680	0.574468	0.0740741	0.224326
23_12759	0.574468	0.0740741	0.212749
26_1314	0.574468	0.0740741	0.188408
26_6355	0.574468	0.037037	0.159414

Appendix D



hat mutants *hac12*, *hac2*, *ham2*, *haf2* and *hag2* do not have compromised AvrRpt2-dependent HR. Leaves were infiltrated with either *Pto* EV (empty vector) or *Pto(avrRpt2)* at OD_{600}= 0.1. Electrolyte leakage was measured 4-12 hours after inoculation. Statistical significance versus Col-0 at each time point determined by two-tailed t-test, n = 6, ** $P \leq 0.01$, * $P \leq 0.05$. Data shown are representative of 2 independent experiments. Error bars indicate standard error.

Bibliography

Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., and Parker, J. E. Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 95(17):10306–10311, 1998.

Agrios, G. N. Plant pathology. 5th edition. *Department of Plant Pathology. University of Florida. United States of America*, 2005.

Akimoto, K., Katakami, H., Kim, H.-J., Ogawa, E., Sano, C. M., Wada, Y., and Sano, H. Epigenetic inheritance in rice plants. *Annals of botany*, 100(2):205–217, 2007.

Aktar, W., Sengupta, D., and Chowdhury, A. Impact of pesticides use in agriculture: their benefits and hazards. *Interdisciplinary toxicology*, 2(1):1–12, 2009.

Alberts, B. *Molecular biology of the cell*. Garland science, 2017.

Alfano, J. R. and Collmer, A. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.*, 42:385–414, 2004.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., et al. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, 301(5633):653–657, 2003.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 1990.

Alvarez, M. E., Nota, F., and Cambiagno, D. A. Epigenetic control of plant immunity. *Molecular Plant Pathology*, 11(4):563–576, 2010.

Alvarez-Venegas, R., Abdallat, A. A., Guo, M., Alfano, J. R., and Avramova, Z. Epigenetic control of a transcription factor at the cross section of two antagonistic pathways. *Epigenetics*, 2(2):106–113, 2007.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2):195–201, 2006.

Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W.-L., Gomez-Gomez, L., Boller, T., Ausubel, F. M., and Sheen, J. MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature*, 415(6875):977–983, 2002.

Avram, S., Funar-Timofei, S., Borota, A., Chennamaneni, S. R., Manchala, A. K., and Muresan, S. Quantitative estimation of pesticide-likeness for agrochemical discovery. *Journal of cheminformatics*, 6(1):42, 2014.

Axtell, M. J. and Staskawicz, B. J. Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell*, 112 (3):369–377, 2003.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37(suppl_2):W202–W208, 2009.

Barnes, P. J. Targeting the epigenome in the treatment of asthma and chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 6 (8):693–696, 2009.

Benhamed, M., Bertrand, C., Servet, C., and Zhou, D.-X. *Arabidopsis* GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. *The Plant Cell*, 18(11):2893–2903, 2006.

Benhamed, M., Martin-Magniette, M.-L., Taconnat, L., Bitton, F., Servet, C., De Clercq, R., De Meyer, B., Buysschaert, C., Rombauts, S., Villarroel, R., et al. Genome-scale *Arabidopsis* promoter array identifies targets of the histone acetyltransferase GCN5. *The Plant Journal*, 56(3):493–504, 2008.

Berger, S. L. The complex language of chromatin regulation during transcription. *Nature*, 447(7143):407–412, 2007.

Berndsen, C. E., Albaugh, B. N., Tan, S., and Denu, J. M. Catalytic mechanism of a MYST family histone acetyltransferase. *Biochemistry*, 46(3):623–629, 2007.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank. *The FEBS Journal*, 80(2):319–324, 1977.

Bertrand, C., Benhamed, M., Li, Y.-F., Ayadi, M., Lemonnier, G., Renou, J.-P., Delarue, M., and Zhou, D.-X. *Arabidopsis* HAF2 gene encoding TATA-binding protein (TBP)-associated factor TAF1, is required to integrate light signals to regulate gene expression and growth. *Journal of Biological Chemistry*, 2004.

Birkenbihl, R. P., Kracher, B., and Somssich, I. E. Induced Genome-Wide Binding of Three *Arabidopsis* WRKY Transcription Factors during Early MAMP-Triggered Immunity. *The Plant Cell*, pages 20–38, 2016.

Blundell, T., Sibanda, B., Sternberg, M., and Thornton, J. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*, 326(6111): 347–352, 1987.

Boller, T. and Felix, G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual review of plant biology*, 60:379–406, 2009.

Bowling, S. A., Clarke, J. D., Liu, Y., Klessig, D. F., and Dong, X. The cpr5 mutant of *Arabidopsis* expresses both NPR1-dependent and NPR1-independent resistance. *The plant cell*, 9(9):1573–1584, 1997.

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., and Leung, H. Plant–pathogen interactions: disease resistance in modern agriculture. *Trends in genetics*, 29(4):233–240, 2013.

Brown, J. K. Yield penalties of disease resistance in crops. *Current opinion in plant biology*, 5(4):339–344, 2002.

Büttner, D. and He, S. Y. Type III protein secretion in plant pathogenic bacteria. *Plant physiology*, 150(4):1656–1664, 2009.

Campi, M., DAndrea, L., Emiliani, J., and Casati, P. Participation of chromatin-remodeling proteins in the repair of ultraviolet-B-damaged DNA. *Plant physiology*, 158(2):981–995, 2012.

Chan, S. W.-L., Zilberman, D., Xie, Z., Johansen, L. K., Carrington, J. C., and Jacobsen, S. E. RNA silencing genes control de novo DNA methylation. *Science*, 303(5662):1336–1336, 2004.

Chen, C., Li, C., Wang, Y., Renaud, J., Tian, G., Kambhampati, S., Saatian, B., Nguyen, V., Hannoufa, A., Marsolais, F., et al. Cytosolic acetyl-CoA promotes histone acetylation predominantly at H3K27 in *Arabidopsis*. *Nature plants*, 3(10): 814, 2017.

Chen, H. and Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics*, 12(1):35, 2011.

Chen, W.-Q., Li, D.-X., Zhao, F., Xu, Z.-H., and Bai, S.-N. One additional histone deacetylase and 2 histone acetyltransferases are involved in cellular patterning of *Arabidopsis* root epidermis. *Plant signaling & behavior*, 11(2):e1131373, 2016.

Chen, Z., Agnew, J. L., Cohen, J. D., He, P., Shan, L., Sheen, J., and Kunkel, B. N. *Pseudomonas syringae* type III effector AvrRpt2 alters *Arabidopsis thaliana* auxin physiology. *Proceedings of the National Academy of Sciences*, 104(50):20131–20136, 2007.

Cheng, C., Gao, X., Feng, B., Sheen, J., Shan, L., and He, P. Plant immune response to pathogens differs with changing temperatures. *Nature communications*, 4:2530–2530, 2013.

Choura, M. Unraveling the WRKY transcription factors network in *Arabidopsis thaliana* by integrative approach. *Network Biology*, 5(2):55, 2015.

Cigliano, R. A., Sanseverino, W., Cremona, G., Ercolano, M. R., Conicella, C., and Consiglio, F. M. Genome-wide analysis of histone modifiers in tomato: gaining an insight into their developmental roles. *BMC genomics*, 14(1):57, 2013.

Clapier, C. R. and Cairns, B. R. The biology of chromatin remodeling complexes. *Annual review of biochemistry*, 78:273–304, 2009.

Clark, D. J. and Kimura, T. Electrostatic mechanism of chromatin folding. *Journal of molecular biology*, 211(4):883–896, 1990.

Clore, G. M., Brünger, A. T., Karplus, M., and Gronenborn, A. M. Application of molecular dynamics with interproton distance restraints to three-dimensional

protein structure determination: a model study of crambin. *Journal of molecular biology*, 191(3):523–551, 1986.

Clough, S. J. and Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The plant journal*, 16(6):735–743, 1998.

Coffey, K., Blackburn, T. J., Cook, S., Golding, B. T., Griffin, R. J., Hardcastle, I. R., Hewitt, L., Huberman, K., McNeill, H. V., Newell, D. R., et al. Characterisation of a Tip60 specific inhibitor, NU9056, in prostate cancer. *PloS one*, 7(10): e45539, 2012.

Consortium, U. Uniprot: the universal protein knowledgebase. *Nucleic acids research*, 45(D1):D158–D169, 2016.

Crick, F. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

Cui, F., Wu, S., Sun, W., Coaker, G., Kunkel, B., He, P., and Shan, L. The *Pseudomonas syringae* type III effector AvrRpt2 promotes pathogen virulence via stimulating *Arabidopsis* auxin/indole acetic acid protein turnover. *Plant physiology*, 162(2):1018–1029, 2013.

Cunnac, S., Lindeberg, M., and Collmer, A. *Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions. *Current opinion in microbiology*, 12(1):53–60, 2009.

da Rosa, J. L., Bajaj, V., Spoonamore, J., and Kaufman, P. D. A small molecule inhibitor of fungal histone acetyltransferase Rtt109. *Bioorganic & medicinal chemistry letters*, 23(10):2853–2859, 2013.

de Torres Zabala, M., Littlejohn, G., Jayaraman, S., Studholme, D., Bailey, T., Lawson, T., Tillich, M., Licht, D., Blter, B., Delfino, L., Truman, W., Mansfield, J., Smirnoff, N., and Grant, M. Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. *Nature Plants*, 1:15074–11, 06 2015.

DeFraia, C. T., Wang, Y., Yao, J., and Mou, Z. Elongator subunit 3 positively regulates plant immunity through its histone acetyltransferase and radical s-adenosylmethionine domains. *BMC plant biology*, 13(1):102, 2013.

Delaney, T. P., Uknes, S., Vernooij, B., Friedrich, L., Weymann, K., Negrotto, D., Gaffney, T., Gut-Rella, M., Kessmann, H., Ward, E., et al. A central role of salicylic acid in plant disease resistance. *Science*, 266(5188):1247–1250, 1994.

Denby, K. J., Kumar, P., and Kliebenstein, D. J. Identification of *Botrytis cinerea* susceptibility loci in *Arabidopsis thaliana*. *The Plant Journal*, 38(3):473–486, 2004.

Denoux, C., Galletti, R., Mammarella, N., Gopalan, S., Werck, D., De Lorenzo, G., Ferrari, S., Ausubel, F. M., and Dewdney, J. Activation of defense response pathways by OGs and Flg22 elicitors in *Arabidopsis* seedlings. *Molecular plant*, 1 (3):423–445, 2008.

Dietrich, R. A., Delaney, T. P., Uknes, S. J., Ward, E. R., Ryals, J. A., and Dangl, J. L. *Arabidopsis* mutants simulating disease resistance response. *Cell*, 77(4): 565–577, 1994.

Ding, B., del Rosario Bellizzi, M., Ning, Y., Meyers, B. C., and Wang, G.-L. HDT701, a histone H4 deacetylase, negatively regulates plant innate immunity by modulating histone H4 acetylation of defense-related genes in rice. *The Plant Cell*, 24(9):3783–3794, 2012.

Dodds, P. N. and Rathjen, J. P. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, 11(8):539–548, 2010.

Dorn, M., e Silva, M. B., Buriol, L. S., and Lamb, L. C. Three-dimensional protein structure prediction: Methods and computational strategies. *Computational biology and chemistry*, 53:251–276, 2014.

Durrant, W. E. and Dong, X. Systemic acquired resistance. *Annu. Rev. Phytopathol.*, 42:185–209, 2004.

Earley, K. W., Shook, M. S., Brower-Toland, B., Hicks, L., and Pikaard, C. S. *In vitro* specificities of *Arabidopsis* co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. *The Plant Journal*, 52(4): 615–626, 2007.

Eulgem, T. and Somssich, I. E. Networks of WRKY transcription factors in defense signaling. *Current opinion in plant biology*, 10(4):366–371, 2007.

Fang, H., Liu, X., Thorn, G., Duan, J., and Tian, L. Expression analysis of histone acetyltransferases in rice under drought stress. *Biochemical and biophysical research communications*, 443(2):400–405, 2014.

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.-M., Gimenez-Ibanez, S., Geerinck, J., Eeckhout, D., Schweizer, F., Godoy, M., Franco-Zorrilla, J. M., et al. The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *The Plant Cell*, 23(2):701–715, 2011.

Fina, J. P. and Casati, P. HAG3, a histone acetyltransferase, affects UV-B responses by negatively regulating the expression of DNA repair enzymes and sunscreen content in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 56(7):1388–1400, 2015.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., et al. Interpro in 2017—beyond protein family and domain annotations. *Nucleic acids research*, 45(D1): D190–D199, 2016.

Flor, H. H. Current status of the gene-for-gene concept. *Annual review of phytopathology*, 9(1):275–296, 1971.

Forli, S., Huey, R., Pique, M. E., Sanner, M., Goodsell, D. S., and Olson, A. J. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nature protocols*, 11(5):905, 2016.

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences*, 111(6):2367–2372, 2014.

Fransz, P. F. and de Jong, J. H. Chromatin dynamics in plants. *Current opinion in plant biology*, 5(6):560–567, 2002.

Frei dit Frey, N., Garcia, A., Bigeard, J., Zaag, R., Bueso, E., Garmier, M., Pateyron, S., de Tauzia-Moreau, M.-L., Brunaud, V., Balzergue, S., et al. Functional analysis of *Arabidopsis* immune-related MAPKs uncovers a role for MPK3 as negative regulator of inducible defences. *Genome biology*, 2014.

Fu, Z. Q. and Dong, X. Systemic acquired resistance: turning local infection into global defense. *Annual review of plant biology*, 64:839–863, 2013.

Fujisawa, T. and Filippakopoulos, P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nature Reviews Molecular Cell Biology*, 2017.

Gandy, M. N., Corral, M. G., Mylne, J. S., and Stubbs, K. A. An interactive database to explore herbicide physicochemical properties. *Organic & biomolecular chemistry*, 13(20):5586–5590, 2015.

Gao, C., Bourke, E., Scobie, M., Famme, M. A., Koolmeister, T., Helleday, T., Eriksson, L. A., Lowndes, N. F., and Brown, J. A. Rational design and validation of a Tip60 histone acetyltransferase inhibitor. *Scientific reports*, 4:5372, 2014.

Gao, X., Chen, X., Lin, W., Chen, S., Lu, D., Niu, Y., Li, L., Cheng, C., McCormack, M., Sheen, J., et al. Bifurcation of *Arabidopsis* NLR immune signaling via Ca2+-dependent protein kinases. *PLoS pathogens*, 9(1):e1003127, 2013.

Gassmann, W., Hinsch, M. E., and Staskawicz, B. J. The *Arabidopsis* RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *The Plant Journal*, 20(3):265–277, 1999.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., and Foley, J. A. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences*, 107(38):16732–16737, 2010.

Gimenez-Ibanez, S., Hann, D. R., Ntoukakis, V., Petutschnig, E., Lipka, V., and Rathjen, J. P. AvrPtoB targets the LysM receptor kinase CERK1 to promote bacterial virulence on plants. *Current biology*, 19(5):423–429, 2009.

Gioia, D., Bertazzo, M., Recanatini, M., Masetti, M., and Cavalli, A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. *Molecules*, 22(11): 2029, 2017.

Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.*, 43:205–227, 2005.

Glickmann, E., Gardan, L., Jacquet, S., Hussain, S., Elasri, M., Petit, A., and Dessaux, Y. Auxin production is a common feature of most pathovars of *Pseudomonas syringae*. *Molecular plant-microbe interactions*, 11(2):156–162, 1998.

Göhre, V., Spallek, T., Häweker, H., Mersmann, S., Mentzel, T., Boller, T., de Torres, M., Mansfield, J. W., and Robatzek, S. Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB. *Current Biology*, 18(23):1824–1832, 2008.

Gómez-Gómez, L. and Boller, T. FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular cell*, 5 (6):1003–1011, 2000.

Goodman, R. N., Novacky, A. J., et al. *The hypersensitive reaction in plants to pathogens: a resistance phenomenon.* 1994.

Greer, J. Comparative model-building of the mammalian serine proteases. *Journal of molecular biology*, 153(4):1027–1042, 1981.

Grienenberger, A., Miotto, B., Sagnier, T., Cavalli, G., Schramke, V., Geli, V., Mariol, M.-C., Berenger, H., Graba, Y., and Pradel, J. The MYST domain acetyltransferase Chameau functions in epigenetic mechanisms of transcriptional repression. *Current biology*, 12(9):762–766, 2002.

Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature*, 389(6649):349–352, 1997.

Gu, Y., Zebell, S. G., Liang, Z., Wang, S., Kang, B.-H., and Dong, X. Nuclear pore permeabilization is a convergent signaling event in effector-triggered immunity. *Cell*, 166(6):1526–1538, 2016.

Hao, G.-F., Wang, F., Li, H., Zhu, X.-L., Yang, W.-C., Huang, L.-S., Wu, J.-W., Berry, E. A., and Yang, G.-F. Computational discovery of picomolar q site inhibitors of cytochrome bc 1 complex. *Journal of the American Chemical Society*, 134(27):11168–11176, 2012.

Hao, G.-F., Zuo, Y., Yang, S.-G., Chen, Q., Zhang, Y., Yin, C.-Y., Niu, C.-W., Xi, Z., and Yang, G.-F. Computational Discovery of Potent and Bioselective Protoporphyrinogen IX Oxidase Inhibitor via Fragment Deconstruction Analysis. *Journal of Agricultural and Food Chemistry*, 65(28):5581–5588, 2017.

Haverkort, A., Boonekamp, P., Hutten, R., Jacobsen, E., Lotz, L., Kessel, G., Visser, R., and Van der Vossen, E. Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. *Potato research*, 51(1):47–57, 2008.

He, P., Shan, L., Lin, N.-C., Martin, G. B., Kemmerling, B., Nürnberger, T., and Sheen, J. Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in *Arabidopsis* innate immunity. *Cell*, 125(3):563–575, 2006.

Heath, M. C. Nonhost resistance and nonspecific plant defenses. *Current opinion in plant biology*, 3(4):315–319, 2000.

Heery, D. M. and Fischer, P. M. Pharmacological targeting of lysine acetyltransferases in human disease: a progress report. *Drug discovery today*, 12(1):88–99, 2007.

Heidel, A. J., Clarke, J. D., Antonovics, J., and Dong, X. Fitness costs of mutations affecting the systemic acquired resistance pathway in *Arabidopsis thaliana*. *Genetics*, 168(4):2197–2206, 2004.

Heil, M., Hilpert, A., Kaiser, W., and Linsenmair, K. E. Reduced growth and seed set following chemical induction of pathogen defence: does systemic acquired resistance (SAR) incur allocation costs? *Journal of Ecology*, 88(4):645–654, 2000.

Hillisch, A., Pineda, L. F., and Hilgenfeld, R. Utility of homology models in the drug discovery process. *Drug discovery today*, 9(15):659–669, 2004.

Holton, N., Nekrasov, V., Ronald, P. C., and Zipfel, C. The phylogenetically-related pattern recognition receptors EFR and XA21 recruit similar immune signaling components in monocots and dicots. *PLoS pathogens*, 11(1):e1004602, 2015.

Hsieh, T.-F. and Fischer, R. L. Biology of chromatin dynamics. *Annu. Rev. Plant Biol.*, 56:327–351, 2005.

Hu, Y., Dong, Q., and Yu, D. *Arabidopsis* WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen *Pseudomonas syringae*. *Plant Science*, 185:288–297, 2012.

Huang, J., Wan, B., Wu, L., Yang, Y., Dou, Y., and Lei, M. Structural insight into the regulation of MOF in the male-specific lethal complex and the non-specific lethal complex. *Cell research*, 22(6):1078, 2012.

Hung, C.-Y., Aspesi Jr, P., Hunter, M. R., Lomax, A. W., and Perera, I. Y. Phosphoinositide-signaling is one component of a robust plant defense response. *Frontiers in plant science*, 5, 2014.

HwangBo, K., Son, S. H., Lee, J. S., Min, S. R., Ko, S. M., Liu, J. R., Choi, D., and Jeong, W. J. Rapid and simple method for DNA extraction from plant and algal species suitable for PCR amplification using a chelating resin Chelex 100. *Plant Biotechnology Reports*, 4(1):49–52, 2010.

Ihaka, R. and Gentleman, R. R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3):299–314, 1996.

Jaghoori, M. M., Bleijlevens, B., and Olabarriaga, S. D. 1001 ways to run AutoDock Vina for virtual screening. *Journal of computer-aided molecular design*, 30(3): 237–249, 2016.

Jenner, R. G. and Young, R. A. Insights into host responses against pathogens from transcriptional profiling. *Nature Reviews Microbiology*, 3(4):281–294, 2005.

Jerzmanowski, A. and Archacki, R. Universal and Lineage-Specific Properties of Linker Histones and SWI/SNF-Chromatin Remodeling Complexes in Plants. In *Plant Epigenetics*, pages 463–492. Springer, 2017.

Jin, Q., Thilmony, R., Zwiesler-Vollick, J., and He, S.-Y. Type III protein secretion in *Pseudomonas syringae*. *Microbes and Infection*, 5(4):301–310, 2003.

Johnson, A. A., Sarthi, J., Pirooznia, S. K., Reube, W., and Elefant, F. Increasing Tip60 HAT levels rescues axonal transport defects and associated behavioral phenotypes in a *Drosophila* Alzheimer's disease model. *Journal of Neuroscience*, 33(17):7535–7547, 2013.

Jones, J. D. and Dangl, J. L. The plant immune system. *Nature*, 444(7117):323–329, 2006.

Jones, J. D., Vance, R. E., and Dangl, J. L. Intracellular innate immune surveillance devices in plants and animals. *Science*, 354(6316):aaf6395, 2016.

Kaimori, J.-Y., Maehara, K., Hayashi-Takanaka, Y., Harada, A., Fukuda, M., Yamamoto, S., Ichimaru, N., Umehara, T., Yokoyama, S., Matsuda, R., et al. Histone H4 lysine 20 acetylation is associated with gene repression in human cells. *Scientific reports*, 6, 2016.

Kaldor, S. W., Kalish, V. J., Davies, J. F., Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., et al.

Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *Journal of medicinal chemistry*, 40(24):3979–3985, 1997.

Katagiri, F., Thilmony, R., and He, S. Y. The *Arabidopsis thaliana-Pseudomonas syringae* interaction. *The Arabidopsis Book*, page e0039, 2002.

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., Feldblyum, T., Nierman, W., Benito, M. I., Lin, X., et al. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *nature*, 408(6814):796–815, 2000.

Kempel, A., Schädler, M., Chrobock, T., Fischer, M., and van Kleunen, M. Tradeoffs associated with constitutive and induced plant resistance against herbivory. *Proceedings of the National Academy of Sciences*, 108(14):5685–5689, 2011.

Kim, G.-W. and Yang, X.-J. Comprehensive lysine acetylomes emerging from bacteria to humans. *Trends in biochemical sciences*, 36(4):211–220, 2011.

Kim, K.-C., Lai, Z., Fan, B., and Chen, Z. *Arabidopsis* WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. *The Plant Cell*, 20(9):2357–2371, 2008.

Kong, L., Qiu, X., Kang, J., Wang, Y., Chen, H., Huang, J., Qiu, M., Zhao, Y., Kong, G., Ma, Z., et al. A *Phytophthora* effector manipulates host histone acetylation and reprograms defense gene expression to promote infection. *Current Biology*, 27(7):981–991, 2017.

Kornet, N. and Scheres, B. Members of the GCN5 histone acetyltransferase complex regulate PLETHORA-mediated root stem cell niche maintenance and transit amplifying cell proliferation in *Arabidopsis*. *The Plant Cell*, 21(4):1070–1079, 2009.

Kouzarides, T. Chromatin modifications and their function. *Cell*, 128(4):693–705, 2007.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. The N terminus of bacterial elongation factor Tu elicits innate immunity in *Arabidopsis* plants. *The Plant Cell*, 16(12):3496–3507, 2004.

Lacombe, S., Rougon-Cardoso, A., Sherwood, E., Peeters, N., Dahlbeck, D., Van Esse, H. P., Smoker, M., Rallapalli, G., Thomma, B. P., Staskawicz, B., et al. Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nature biotechnology*, 28(4):365–369, 2010.

Laskowski, R. A. and Swindells, M. B. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery, 2011.

Latrasse, D., Benhamed, M., Henry, Y., Domenichini, S., Kim, W., Zhou, D.-X., and Delarue, M. The MYST histone acetyltransferases are essential for gametophyte development in *Arabidopsis*. *BMC plant biology*, 8(1):121, 2008.

Latrasse, D., Jégu, T., Li, H., de Zelicourt, A., Raynaud, C., Legras, S., Gust, A., Samajova, O., Veluchamy, A., Rayapuram, N., et al. MAPK-triggered chromatin reprogramming by histone deacetylase in plant innate immunity. *Genome biology*, 18(1):131, 2017.

Levitt, M. Accurate modeling of protein conformation by automatic segment matching. *Journal of molecular biology*, 226(2):507–533, 1992.

Lewis, L. A., Polanski, K., de Torres-Zabala, M., Jayaraman, S., Bowden, L., Moore, J., Penfold, C. A., Jenkins, D. J., Hill, C., Baxter, L., et al. Transcriptional dynamics driving MAMP-triggered immunity and pathogen effector-mediated immunosuppression in *Arabidopsis* leaves following infection with *Pseudomonas syringae pv tomat*o DC3000. *The Plant Cell*, 27(11):3038–3064, 2015.

Li, C., Xu, J., Li, J., Li, Q., and Yang, H. Involvement of *Arabidopsis* HAC family genes in pleiotropic developmental processes. *Plant signaling & behavior*, 9(3): 426–35, 2014a.

Li, G., Liu, S., Wang, J., He, J., Huang, H., Zhang, Y., and Xu, L. ISWI proteins participate in the genome-wide nucleosome distribution in *Arabidopsis*. *The Plant Journal*, 78(4):706–714, 2014b.

Lindell, S. D., Pattenden, L. C., and Shannon, J. Combinatorial chemistry in the agrosciences. *Bioorganic & medicinal chemistry*, 17(12):4035–4046, 2009.

Liu, X., Yang, S., Zhao, M., Luo, M., Yu, C.-W., Chen, C.-Y., Tai, R., and Wu, K. Transcriptional repression by histone deacetylases in plants. *Molecular plant*, 7 (5):764–772, 2014.

Long, J. A., Ohno, C., Smith, Z. R., and Meyerowitz, E. M. TOPLESS regulates apical embryonic fate in *Arabidopsis*. *Science*, 312(5779):1520–1523, 2006.

López-Ramos, M. and Perruccio, F. HPPD: ligand-and target-based virtual screening on a herbicide target. *Journal of chemical information and modeling*, 50(5): 801–814, 2010.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. Structure validation by cα geometry: ϕ, ψ and cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3):437–450, 2003.

Lu, F., Wang, H., Wang, S., Jiang, W., Shan, C., Li, B., Yang, J., Zhang, S., and Sun, W. Enhancement of innate immune system in monocot rice by transferring the dicotyledonous elongation factor Tu receptor EFR. *Journal of integrative plant biology*, 57(7):641–652, 2015.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389 (6648):251–260, 1997.

Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. An overview of the structures of protein-DNA complexes. *Genome biology*, 1(1):reviews001–1, 2000.

Macek, B., Waanders, L. F., Olsen, J. V., and Mann, M. Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, 5(5):949–958, 2006.

Macho, A. P. and Zipfel, C. Plant PRRs and the activation of innate immune signaling. *Molecular cell*, 54(2):263–272, 2014.

MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616, 1998.

Mackey, D., Holt, B. F., Wiig, A., and Dangl, J. L. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell*, 108(6):743–754, 2002.

Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R., and Dangl, J. L. *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell*, 112(3):379–389, 2003.

Maekawa, T., Kracher, B., Vernaldi, S., van Themaat, E. V. L., and Schulze-Lefert, P. Conservation of NLR-triggered immunity across plant lineages. *Proceedings of the National Academy of Sciences*, 109(49):20119–20123, 2012.

Maere, S., Heymans, K., and Kuiper, M. BiNGO: a Cytoscape plugin to assess over-representation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448–3449, 2005.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S. V., Machado, M. A., et al. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular plant pathology*, 13(6):614–629, 2012.

Mao, Y., Pavangadkar, K. A., Thomashow, M. F., and Triezenberg, S. J. Physical and functional interactions of *Arabidopsis* ADA2 transcriptional coactivator proteins with the acetyltransferase GCN5 and with the cold-induced transcription factor CBF1. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1759(1):69–79, 2006.

McLeay, R. C. and Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC bioinformatics*, 11(1):165, 2010.

Meldau, S., Ullman-Zeunert, L., Govind, G., Bartram, S., and Baldwin, I. T. MAPK-dependent JA and SA signalling in *Nicotiana attenuata* affects plant growth and fitness during competition with conspecifics. *BMC Plant Biology*, 12(1):213, 2012.

Menke, F. L., Van Pelt, J. A., Pieterse, C. M., and Klessig, D. F. Silencing of the mitogen-activated protein kinase MPK6 compromises disease resistance in *Arabidopsis*. *The Plant Cell*, 16(4):897–907, 2004.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. Genome-wide analysis of NBS-LRR–encoding genes in *Arabidopsis*. *The Plant Cell*, 15(4): 809–834, 2003.

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H., and Shibuya, N. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 104(49):19613–19618, 2007.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16): 2785–2791, 2009.

Moult, J., Fidelis, K., Zemla, A., and Hubbard, T. Critical assessment of methods of protein structure prediction (CASP)-round V. *Proteins: Structure, Function, and Bioinformatics*, 53(S6):334–339, 2003.

Narusaka, M., Kubo, Y., Hatakeyama, K., Imamura, J., Ezura, H., Nanasato, Y., Tabei, Y., Takano, Y., Shirasu, K., and Narusaka, Y. Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. *PLoS One*, 8(2):e55954, 2013.

Navarro, L., Zipfel, C., Rowland, O., Keller, I., Robatzek, S., Boller, T., and Jones, J. D. The transcriptional innate immune response to flg22. Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis. *Plant physiology*, 135(2):1113–1128, 2004.

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O., and Jones, J. D. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science*, 312(5772):436–439, 2006.

Needleman, S. B. and Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, 1970.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):33, 2011.

of the United Nations, F. A. O. FAOSTAT Statistics Database, 2001. Data retrieved from FAOSTAT Statistics Database, `http://www.fao.org/land-water/databases-and-software/crop-information/tomato/en/`.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, 44(D1):D733–D745, 2015.

Oliveros, J. C. VENNY. An interactive tool for comparing lists with Venn Diagrams. *http://bioinfogp. cnb. csic. es/tools/venny/index. html*, 2007.

O'Malley, R. C., Huang, S.-s. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., and Ecker, J. R. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, 165(5):1280–1292, 2016.

Pandey, R., MuÈller, A., Napoli, C. A., Selinger, D. A., Pikaard, C. S., Richards, E. J., Bender, J., Mount, D. W., and Jorgensen, R. A. Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic acids research*, 30(23):5036–5055, 2002.

Pearson, W. R. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in enzymology*, 183:63–98, 1990.

Pieterse, C. M., Leon-Reyes, A., Van der Ent, S., and Van Wees, S. C. Networking by small-molecule hormones in plant immunity. *Nature chemical biology*, 5(5): 308–316, 2009.

Popp, J., Pető, K., and Nagy, J. Pesticide productivity and food security: a review. *Agronomy for sustainable development*, 33(1):243–255, 2013.

Poulios, S. and Vlachonasios, K. E. Synergistic action of histone acetyltransferase GCN5 and receptor CLAVATA1 negatively affects ethylene responses in *Arabidopsis thaliana*. *Journal of experimental botany*, 67(3):905–918, 2015.

Raevsky, A., Sharifi, M., Samofalova, D., Karpov, P., and Blume, Y. 3D structure prediction of histone acetyltransferase proteins of the MYST family and their interactome in *Arabidopsis thaliana*. *Journal of molecular modeling*, 22(11):256, 2016.

Ren, X., Chen, Z., Liu, Y., Zhang, H., Zhang, M., Liu, Q., Hong, X., Zhu, J.-K., and Gong, Z. ABO3, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in *Arabidopsis*. *The Plant Journal*, 63(3): 417–429, 2010.

Reuber, T. L. and Ausubel, F. M. Isolation of *Arabidopsis* genes that differentiate between resistance responses mediated by the RPS2 and RPM1 disease resistance genes. *The Plant Cell*, 8(2):241–249, 1996.

Rice, P., Longden, I., and Bleasby, A. EMBOSS: the European molecular biology open software suite, 2000.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. LIMMA powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.

Ritter, C. and Dangl, J. L. Interference between two specific pathogen recognition events mediated by distinct plant disease resistance genes. *The Plant Cell*, 8(2): 251–257, 1996.

Robert-Seilaniantz, A., Grant, M., and Jones, J. D. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annual review of phytopathology*, 49:317–343, 2011.

Rognan, D. Docking methods for virtual screening: principles and recent advances. *Virtual Screening: Principles, Challenges, and Practical Guidelines*, pages 153–176, 2011.

Rost, B. Twilight zone of protein sequence alignments. *Protein engineering*, 12(2): 85–94, 1999.

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., et al. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *The EMBO journal*, 30(10):1928–1938, 2011.

Roush, R. and Tabashnik, B. E. *Pesticide resistance in arthropods*. Springer Science & Business Media, 2012.

Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. WRKY transcription factors. *Trends in plant science*, 15(5):247–258, 2010.

Sali, A. and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3):779–815, 1993.

Sander, T., Freyss, J., von Korff, M., and Rufener, C. DataWarrior: an open-source program for chemistry aware data visualization and analysis. *Journal of chemical information and modeling*, 55(2):460–473, 2015.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7):671–675, 2012.

Schoonbeek, H.-j., Wang, H.-H., Stefanato, F. L., Craze, M., Bowden, S., Wallington, E., Zipfel, C., and Ridout, C. J. *Arabidopsis* EF-Tu receptor enhances bacterial disease resistance in transgenic wheat. *New Phytologist*, 206(2):606–613, 2015.

Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

Schulze, B., Mentzel, T., Jehle, A. K., Mueller, K., Beeler, S., Boller, T., Felix, G., and Chinchilla, D. Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. *Journal of Biological Chemistry*, 285(13):9444–9451, 2010.

Schwessinger, B., Bahar, O., Thomas, N., Holton, N., Nekrasov, V., Ruan, D., Canlas, P. E., Daudi, A., Petzold, C. J., Singan, V. R., et al. Transgenic expression of the dicotyledonous pattern recognition receptor EFR in rice leads to ligand-dependent activation of defense responses. *PLoS pathogens*, 11(3):e1004809, 2015.

Scofield, S. R., Tobias, C. M., Rathjen, J. P., Chang, J. H., Lavelle, D. T., Michelmore, R. W., and Staskawicz, B. J. Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science*, 274(5295):2063–2065, 1996.

Sebastian, J. and Lee, J.-Y. Root apical meristems. *eLS*, 2013.

Servet, C., e Silva, N. C., and Zhou, D.-X. Histone acetyltransferase AtGCN5/HAG1 is a versatile regulator of developmental and inducible gene expression in *Arabidopsis*. *Molecular plant*, 3(4):670–677, 2010.

Shan, L., He, P., Li, J., Heese, A., Peck, S. C., Nürnberger, T., Martin, G. B., and Sheen, J. Bacterial effectors target the common signaling partner BAK1 to disrupt multiple MAMP receptor-signaling complexes and impede plant immunity. *Cell host & microbe*, 4(1):17–27, 2008.

Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J. E., and Innes, R. W. Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science*, 301(5637): 1230–1233, 2003.

Sheard, L. B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T. R., Kobayashi, Y., Hsu, F.-F., Sharon, M., He, S. Y., et al. Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature*, 468(7322):400–405, 2010.

Shen, M.-y. and Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein science*, 15(11):2507–2524, 2006.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1):539, 2011.

Singh, J., Chuaqui, C. E., Boriack-Sjodin, P. A., Lee, W.-C., Pontz, T., Corbley, M. J., Cheung, H.-K., Arduini, R. M., Mead, J. N., Newman, M. N., et al. Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFβ receptor kinase (TβRI). *Bioorganic & medicinal chemistry letters*, 13(24):4355–4359, 2003.

Singh, P., Yekondi, S., Chen, P.-W., Tsai, C.-H., Yu, C.-W., Wu, K., and Zimmerli, L. Environmental History Modulates *Arabidopsis* Pattern-Triggered Immunity in a HISTONE ACETYLTRANSFERASE1–Dependent Manner. *The Plant Cell*, 26 (6):2676–2688, 2014.

Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4):355–362, 1993.

Sippl, M. J. Knowledge-based potentials for proteins. *Current opinion in structural biology*, 5(2):229–235, 1995.

Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. Computational methods in drug discovery. *Pharmacological reviews*, 66(1):334–395, 2014.

Smith, T. F. and Waterman, M. S. Identification of common molecular subsequences. *Journal of molecular biology*, 147(1):195–197, 1981.

Sousa, S. F., Ribeiro, A. J., Coimbra, J., Neves, R., Martins, S., Moorthy, N., Fernandes, P., and Ramos, M. Protein-ligand docking in the new millennium–a retrospective of 10 years in the field. *Current medicinal chemistry*, 20(18):2296–2314, 2013.

Spoel, S. H., Koornneef, A., Claessens, S. M., Korzelius, J. P., Van Pelt, J. A., Mueller, M. J., Buchala, A. J., Métraux, J.-P., Brown, R., Kazan, K., et al. NPR1 modulates cross-talk between salicylate-and jasmonate-dependent defense pathways through a novel function in the cytosol. *The Plant Cell*, 15(3):760–770, 2003.

Springer, N. M. Epigenetics and crop improvement. *Trends in genetics*, 29(4): 241–247, 2013.

Springer, N. M. and Schmitz, R. J. Exploiting induced and natural epigenetic variation for crop improvement. *Nature reviews genetics*, 18(9):563–575, 2017.

Thomas, T. and Voss, A. K. The diverse biological roles of MYST histone acetyltransferase family proteins. *Cell Cycle*, 6(6):696–704, 2007.

Tice, C. M. Selecting the right compounds for screening: does Lipinski's Rule of 5 for pharmaceuticals apply to agrochemicals? *Pest management science*, 57(1): 3–16, 2001.

Trievel, R. C., Rojas, J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., Allis, C. D., Berger, S. L., and Marmorstein, R. Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *Proceedings of the National Academy of Sciences*, 96(16):8931–8936, 1999.

Trott, O. and Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

Tsuda, K. and Katagiri, F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current opinion in plant biology*, 13 (4):459–465, 2010.

UN. World population prospects: the 2012 revision. *Population division of the department of economic and social affairs of the United Nations Secretariat, New York*, 2013.

Uzunova, V. V., Quareshy, M., Del Genio, C. I., and Napier, R. M. Tomographic docking suggests the mechanism of auxin receptor TIR1 selectivity. *Open biology*, 6(10):160139, 2016.

Varghese, J. N. Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug development research*, 46(3-4):176–196, 1999.

Vinatzer, B. A., Teitzel, G. M., Lee, M.-W., Jelenska, J., Hotton, S., Fairfax, K., Jenrette, J., and Greenberg, J. T. The type III effector repertoire of *Pseudomonas syringae pv. syringae* B728a and its role in survival and disease on host and non-host plants. *Molecular microbiology*, 62(1):26–44, 2006.

Vlachonasios, K. E., Thomashow, M. F., and Triezenberg, S. J. Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect

Arabidopsis growth, development, and gene expression. *The Plant Cell*, 15(3): 626–638, 2003.

Wang, S., Gu, Y., Zebell, S. G., Anderson, L. K., Wang, W., Mohan, R., and Dong, X. A noncanonical role for the CKI-RB-E2F cell-cycle signaling pathway in plant effector-triggered immunity. *Cell host & microbe*, 16(6):787–794, 2014.

Wang, T., Xing, J., Liu, X., Liu, Z., Yao, Y., Hu, Z., Peng, H., Xin, M., Zhou, D.-X., Zhang, Y., et al. Histone acetyltransferase general control non-repressed protein 5 (GCN5) affects the fatty acid composition of *Arabidopsis thaliana* seeds by acetylating fatty acid desaturase3 (FAD3). *The Plant Journal*, 88(5):794–808, 2016.

Wang, Y., An, C., Zhang, X., Yao, J., Zhang, Y., Sun, Y., Yu, F., Amador, D. M., and Mou, Z. The *Arabidopsis* elongator complex subunit2 epigenetically regulates plant immune responses. *The Plant Cell*, 25(2):762–776, 2013.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191, 2009.

Webb, B. and Sali, A. Protein structure modeling with MODELLER. In *Functional Genomics*, pages 39–54. Springer, 2017.

Wegel, E. and Shaw, P. Gene activation and deactivation related changes in the three-dimensional structure of chromatin. *Chromosoma*, 114(5):331–337, 2005.

Wei, W., Plovanich-Jones, A., Deng, W.-L., Jin, Q.-L., Collmer, A., Huang, H.-C., and He, S. Y. The gene coding for the Hrp pilus structural protein is required for type III secretion of Hrp and Avr proteins in *Pseudomonas syringae pv. tomato*. *Proceedings of the National Academy of Sciences*, 97(5):2247–2252, 2000.

Wiederstein, M. and Sippl, M. J. Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, 35(suppl_2):W407–W410, 2007.

Wildermuth, M. C., Dewdney, J., Wu, G., and Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature*, 414(6863): 562–565, 2001.

Windram, O., Madhou, P., McHattie, S., Hill, C., Hickman, R., Cooke, E., Jenkins, D. J., Penfold, C. A., Baxter, L., Breeze, E., et al. *Arabidopsis* defense against *Botrytis cinerea*: chronology and regulation deciphered by high-resolution temporal transcriptomic analysis. *The Plant Cell Online*, 24(9):3530–3557, 2012.

Wolber, G. and Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of chemical information and modeling*, 45(1):160–169, 2005.

Xiao, H., Chung, J., Kao, H.-Y., and Yang, Y.-C. Tip60 is a co-repressor for STAT3. *Journal of Biological Chemistry*, 278(13):11197–11204, 2003.

Xiao, J., Zhang, H., Xing, L., Xu, S., Liu, H., Chong, K., and Xu, Y. Requirement of histone acetyltransferases HAM1 and HAM2 for epigenetic modification of FLC in regulating flowering in *Arabidopsis*. *Journal of plant physiology*, 170(4):444–451, 2013.

Xin, X.-F. and He, S. Y. *Pseudomonas syringae pv. tomato* DC3000: a model pathogen for probing disease susceptibility and hormone signaling in plants. *Annual review of phytopathology*, 51:473–498, 2013.

Xu, X., Chen, C., Fan, B., and Chen, Z. Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. *The Plant Cell*, 18(5):1310–1326, 2006.

Yan, Y., Barlev, N. A., Haley, R. H., Berger, S. L., and Marmorstein, R. Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. *Molecular cell*, 6(5):1195–1205, 2000.

Yang, X.-J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic acids research*, 32(3):959–976, 2004.

Yuan, H., Rossetto, D., Mellert, H., Dang, W., Srinivasan, M., Johnson, J., Hodawadekar, S., Ding, E. C., Speicher, K., Abshiru, N., et al. MYST protein acetyltransferase activity requires active site lysine autoacetylation. *The EMBO journal*, 31(1):58–70, 2012.

Zhang, J., Shao, F., Li, Y., Cui, H., Chen, L., Li, H., Zou, Y., Long, C., Lan, L., Chai, J., et al. A *Pseudomonas syringae* effector inactivates MAPKs to suppress PAMP-induced immunity in plants. *Cell host & microbe*, 1(3):175–185, 2007.

Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., and Jacobsen, S. E. Genome-wide analysis of mono-, di-and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome biology*, 10(6):R62, 2009.

Zhang, X., Yao, J., Zhang, Y., Sun, Y., and Mou, Z. The *Arabidopsis* Mediator complex subunits MED14/SWP and MED16/SFR6/IEN1 differentially regulate defense gene expression in plant immune responses. *The Plant Journal*, 75(3): 484–497, 2013.

Zhou, C., Zhang, L., Duan, J., Miki, B., and Wu, K. HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in *Arabidopsis*. *The Plant Cell*, 17(4):1196–1204, 2005.

Zipfel, C. and Felix, G. Plants and animals: a different taste for microbes? *Current opinion in plant biology*, 8(4):353–360, 2005.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E. J., Jones, J. D., Felix, G., and Boller, T. Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature*, 428(6984):764–767, 2004.

Zou, Y. and Bi, X. Positive roles of SAS2 in DNA replication and transcriptional silencing in yeast. *Nucleic acids research*, 36(16):5189–5200, 2008.